41

ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XLI NUMBER 1 MARCH 1995

The Alberta Journal of Educational Research

Published in March, June, September, and December by the Faculty of Education, University of Alberta.

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
Mount Allison University

George Buck
University of Alberta

Ardra L. Cole
Ontario Institute for Studies in Education

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Heather Ryan
Massey University

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Ken Ward
Robert H. Short

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, the Alberta Advisory Committee for Educational Studies, and the Research Grants Office of the University of Alberta

The subscription rate is $32.00 per year for individuals, $45.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Office of the Dean
845 Education Centre South
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0390

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XLI, Number 1, March 1995

Articles

John Biggs **1** Assessing for Learning: Some Dimensions Underlying New Approaches to Educational Assessment

Debora Barnett-Foster
Philip Nagy **18** A Comparison of Undergraduate Test Response Strategies for Multiple-choice and Constructed-response Questions

Jeremy Hull
Ron Phillips
Eleoussa Polyzoi **36** Indian Control and the Delivery of Special Education Services to Students in Band-operated Schools in Manitoba

Grace V. Malicky
Charles A. Norman **63** Perceptions of Literacy and Adult Literacy Programs

Richard Conte
Jac J.W. Andrews
Melanie Loomer
Gillian Hutton **84** A Classroom-based Social Skills Intervention for Children with Learning Disabilities

John A. Ross
J. Bradley Cousins **103** Giving and Receiving Explanations in Cooperative Learning Groups

Book Reviews

Tracey Derwing **122** *Genie: An Abused Child's Flight from Silence* by Russ Rymer

Gregor Wolbring **124** *Violence and Abuse in the Lives of People With Disabilities* by Dick Sobsey

ISSN 0002-4805

Acknowledgment

The following people have helped to maintain the quality of AJER by reviewing manuscripts during the past year. Their contributions are greatly appreciated.

Assheton-Smith, M.
Bain, B.
Baine, D.
Bateson, D.
Beebe, M.
Belke, T.
Benson, G.
Bilash, O.
Blakey, J.
Carson, T.
Chambers, C.
Chapman, J.
Cole, A.
Connors, J.
Dawson, S.
Derwing, T.M.
Donald, J.
Draper, J.
Edwards, J.
Epling, W.F.
Gaskell, J.
Glegg, A.
Goldberg, J.
Golic, J.
Haggerty, S.
Harley, B.
Housego, B.E.J.
Juliebo, J.
Kalish, C.
Kirby, J.
Kirman, J.
Konrad, A.
Krentz, C.
Krull, C.
Kysela, G.M.
LaRocque, L.
Leithwood, K.
LeRoy, C.
Maguire, M.
Maguire, T.O.
Marini, Z.
McCann, S.
McFadden, C.
Mohan, B.
Norris, J.
Norris, S.
Oberg, A.
Owens, D.T.
Pearson, A.
Phillips, L.
Pierce, W.D.
Piper, T.
Ray, D.
Rees, R.
Rogers, T.
Rowell, P.
Ryan, B.
Ryan, H.
Samiroden, W.
Snart, F.
Sobsey, R.J.
Stapleton, J.
Symons, F.
Taafe, R.
Taylor, G.
Thomas, A.
Toohey, K.
Uhl, N.
VanBrunschot, E.
Ward, K.
Wiener, J.
Wilgosh, L.
Wilson, R.J.
Winne, P.
Wyrostok, N.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, 1995, 1-17*

John Biggs
University of Hong Kong

Assessing for Learning: Some Dimensions Underlying New Approaches to Educational Assessment

The theory and practice of assessing learning are currently undergoing a paradigm shift. The critical realization in producing this change is that educational considerations should drive testing, not psychometric or political ones. Three dimensions interact to yield different modes of assessment, including different kinds of performance assessment: the measurement versus the standards model of testing, quantitative and qualitative assumptions as to the nature of what is learned, and whether the learning and testing is situated or decontextualized. The modes of assessment so generated are suited for different educational aims, but the most appropriate modes are underrepresented in current practice, quantitative and decontextualized modes being greatly overrepresented, resulting in backwash often deleterious to teaching and learning. Conceptual and structural difficulties in implementing qualitative and situated modes of assessment are discussed.

La théorie et la pratique de l'évaluation de l'apprentissage subissent présentement un changement de forme draconien. Ce changement de la perspective de l'évaluation de l'apprentissage relèverait du fait que se sont les considérations pédagogiques éducationnelles qui devraient diriger l'orientation du testing et non les orientations psychométriques ou politiques. Trois dimensions s'entrecroisent produisant différents modes d'évaluations incluant différentes sortes d'habiletés et performances: le modèle de testing de la mesure comparé au modèle de testing des standards, les prédispositions qualitatives et quantitatives lorsqu'elles concernent la nature de ce qui est appris, et si l'apprentissage et le testing sont situés ou décontextualisés. Les façons d'évaluation qui en sont par conséquent créées semblent être adaptées à différents buts éducationnels, mais semblerait-il que les méthodes les plus appropriées sont sous-représentées dans la pratique courante de l'enseignement, et que les méthodes quantitatives et décontextualisées semblent être sur-représentées ce qui semble effectuer un certain recul sur l'enseignement et l'apprentissage. On discute des difficultés conceptuelles et structurelles de l'implantation des méthodes d'évaluation qualitatives et situées.

Assumptions Underlying Assessing and Learning

The Measurement Model

For many years, perhaps the greater part of this century, the assessment of academic learning proceeded with little change and with little perceived need for change. Test developers, psychometricians, and teacher educators accepted the *measurement* model of assessment, based on trait theory, which requires that test item scores are sufficiently consistent with each other that testees may be ordered along a single dimension represented by that intra-test consistency (Taylor, 1994). The technologies of test construction, item selection, and establishing reliability and validity followed from these basic assumptions. That

John Biggs is a professor of education in the University of Hong Kong, a post he has held since 1987. His major research area is learning and motivation in institutional settings, and he is particularly interested in the effects of approaches to teaching and assessment on the quality of student learning.

technology became sophisticated and powerful in its appropriate applications. Classroom teachers for their part probably held the accepted framework as espoused theory, their theory-in-use an uneasy compromise between that and established practice (but see Cizek, 1993); when they do try to put their espoused theory directly into practice they usually get it wrong (Marso & Pigge, 1991).

The problem is not teacher incompetence, but the fact that the "technology of assessment that grew out of test theory ... lacked a basis in psychological theory" (Wilson & Kirby, 1994, p. 107); even more to the point, it lacked a basis in educational theory and the knowledge base of teaching (Haertel, 1991). With changing educational philosophies traditional test theory became decreasingly appropriate to the majority of classroom testing occasions. Specifically, the model assumes the stability of the dimension being tested, which is appropriate when the task is to discriminate between individual performances and to predict future performance, but is inappropriate to assess individual or group outcomes in relation to a particular curriculum because the intervention of *teaching* assumes change. As social expectations became increasingly that most students should complete grades K through 12, acquiring basic skills, competences and certain declarative knowledge on the way, selecting students for different levels of schooling ceased to be a major issue, but ensuring that students attained these standards acceptably (however one defines that in the event) was a definite issue of public concern. In an age of electronic learning assessment practices were being driven by steam.

The Standards Model

Models of assessment based on outcomes reaching a predetermined standard have also been around for many years—for example, the notion of a First Class Honors in the British University system—but these ideas were subjective and they specified no technology. The criterion-referenced testing (CRT) movement was probably the first systematic attempt to formulate and enact a set of assumptions critically different from those underlying the measurement model in that instructional and assessment strategy were specifically linked (Bloom, Hasting, & Madaus, 1971; Popham & Husek, 1969). CRT is an example of what Taylor (1994) calls the standards model of assessment, which is based on such assumptions as: public standards can be set; they can be reached by most students, albeit by different kinds of performance; and fair and consistent judgments are possible to determine whether the standards have been met. These assumptions about assessment are isomorphic to those underlying learning and instruction.

The traditional CRT model fell short, however, on the nature of the standards that were set, which in some important respects were no improvement on those underlying the measurement model itself. Although CRT was an improvement in terms of the link between assessment and instruction, the links between these and the question of what is learned remained unchanged. CRT had as limited a view of learning as had the measurement model.

Two basic conceptions of the nature of learning exist in our educational thinking, quantitative and qualitative (Cole, 1990; Marton, dall'Alba, & Beaty, in press). How we view learning will determine how we go about learning, teaching, and, to the present point, assessment.

The Quantitative Tradition

The quantitative tradition has the longest history in educational thinking, stemming from the positivist tradition in the social sciences (Moss, 1992). Learning is conceived as acquiring "specific discrete skills described as precise well-delimited behaviors" (Cole, 1990, p. 2). These contents of learning are treated as discrete quanta of declarative or procedural knowledge; as far as assessment is concerned, any one quantum is treated as functionally independent of any other. In this view, the curriculum becomes in effect a list of discrete units: facts, skills, competences, behavioral objectives, performance indicators, and the like, and assessment a matter of how many of these have been attained.

Teaching is conceived as transmitting knowledge from teacher to learner, and many delivery and assessment systems are based on the transmission model up to college and university level (Trigwell, Prosser, & Taylor, 1994). The teacher's task is to know the subject and expound it clearly, the learner's to receive accurately.

In assessment practice, the contents of knowledge are seen as learned in binary units (correct/incorrect), the correct units being summed to give an aggregate score that yields an index of competence in what is learned. Multiple-choice tests enact this clearly; competence is represented as a total score of all items correct, any one item being worth the same as any other. Lohman (1993) cites an example of a multiple-choice test given to grade 5 children when the 200th anniversary of the United States Constitution was being celebrated. The only item on the test referring to Thomas Jefferson was "Who was the signer of the constitution who had six children?" The problem is that this tells the child that every idea in the test is equally important. Lohman (1993) recounts that a year later he asked a child in this class what she remembered of Thomas Jefferson: the number of his children, but nothing of his role in the Constitution. Such testing tells students that

> There is no need to separate main ideas from details; all are worth one point. And there is no need to assemble these ideas into a coherent summary or to integrate them with anything else because that is not required. (p. 19)

However, particularly in large classes where marking schemes are used, essays are frequently treated in the same manner, a mark being given as each "correct" or "acceptable" point is made, with possibly bonus points for argument or style. I am not saying that individual items in a multiple-choice test may not be clever, demanding, or substantively significant or that the essay question is unsuccessful in eliciting high cognitive-level engagement, but that the treatment of the item scores and test marks assumes their mutual equivalence, independence, and additivity. Students know this and are strategic in exploiting it (Biggs, 1973; Crooks, 1988); in timed examinations, for instance, in view of the law of diminishing returns with respect to mark allocation (time spent on the first half of an essay almost always nets more marks than the same time spent on the second half), attempting all five, say, questions but finishing none will capture more marks than writing a properly structured answer to only four. It is interesting to note that the national model in the United States for setting both objective and essay questions, the Bloom

taxonomy, "has no category for the organization of factual knowledge" (Lohman, 1993, p. 22).

Nevertheless, whether the backwash effects of multiple-choice and other testing formats in the quantitative tradition are seen as deleterious or beneficial for instruction is a much debated question, the answer no doubt depending on whether one holds a quantitative view of learning or not. There have been several warnings of the harmful influences, ranging from the popular (Hoffman, 1962) to the scholarly (Frederiksen, 1984), whereas psychometricians who hold avowed quantitative assumptions about the nature of learning see deliberate and systematic teaching to the test as sound teaching (Shepard, 1991), elevated by Popham (1987) to the strategy of measurement-driven instruction.

It will be noted that quantitative assumptions underlie both norm-referenced testing (NRT) in the traditional measurement model and criterion-referenced testing (CRT) in the Popham and mastery models. The issue is thus not simply the measurement versus the standards models (Taylor, 1994), but a question of one's assumptions about the nature of the learning to be assessed.

The Qualitative Tradition

The qualitative tradition has its roots in 19th-century phenomenology, and later in Gestalt psychology, but it is only recently that it has got as far as offering an alternative paradigm as far as educational decision making on a wide front is concerned. The underlying theory of learning is *constructivism*, a family of theories rather than any one, according to which students are assumed to learn cumulatively, actively interpreting and incorporating new material with what they already know. Different theories variously emphasize the individual, social, cognitive, saccadic, contextual, or emergent natures of learning, but all agree on an active learner seeking meaning by constructing knowledge rather than by receiving and storing knowledge. Poststructuralists emphasize the social construction of knowledge (Delandsheere & Petrovsky, 1994), whereas my own view and that underlying this article stems ultimately from cognitive psychology (Biggs & Moore, 1993). In this latter view understanding changes progressively as people learn, with qualitative changes taking place in the nature both of what is learned and how it is structured. Understanding of a topic thus evolves cumulatively over the long haul, having *horizontal* interconnections with other topics and subjects and *vertical* interconnections with previous and subsequent learnings in the same topic.

As the contents of learning are meanings, the curriculum question is to decide what meanings or levels of understanding are reasonable at the stage of learning in question. The teacher's task is not then to transmit correct understandings, but to help students construct understandings that are progressively more mature and congruent with accepted thinking, recognizing that in many subject areas students' everyday experiences have helped them to construct alternative ways of construing their world. Teaching techniques may be expository at times, but essentially they will involve more effective ways of eliciting constructive activity on the part of the student, such as the deliberate and explicit use of the relevant knowledge base, peer and student-teacher interaction, a motivating context, and much student activity, both reflective or self-directed as well as task-directed (Biggs & Moore, 1993).

Whereas the logic of assessment from a quantitative point of view implies aggregating units of learning taken cross-sectionally with respect to time, that from the qualitative tradition implies charting longitudinal growth over time, from relative ignorance to relative competence: establishing the limits of *relative* is of course a major curriculum question. The outcomes of learning become the constructions the learner has made at any given point in the process. If that growth in competence can be described in recognizable stages, then so much the better because these stages can then become assessment targets (Biggs & Collis, 1989; Clark, Scarino, & Brownell, 1994).

Assessment within the qualitative framework may be of two basic kinds:

1. *developmental*, the purpose of which is to discover where students are in the development of understanding or competence in the domain of the concept or skill in question, the focus here being pure or discipline-based knowledge;
2. *ecological*, the purpose of which is to discover if students can carry out tasks that are "worthwhile, significant, and meaningful" (Archbald & Newman, 1988, p. 1), the focus here being on applications and problem solving.

As far as issues of assessment are concerned, there is some similarity between quantitative/qualitative distinction and Lohman's distinction between crystallized abilities as being involved in teaching and assessing for near transfer in familiar situations, and likewise for fluid abilities for far transfer in unfamiliar situations; the first involve lower-order, and the second, higher-order cognitive skills. Education should involve both, but common assessment tasks evoke predominantly the former (Lohman, 1993).

Situated and Decontextualized Assessment

Ecological assessment thus becomes the qualitative assessment of applied procedural knowledge, and as such is closely related to the more general movement referred to as "authentic" assessment (Newman & Archbald, 1992; Wiggins, 1989), which insists that the context of testing should reflect the goals of learning insofar as they require students to think, decide, and act in the real world (Archbald & Newman, 1988). However, the term *authentic* grabs the moral high ground rather, and *performance* assessment is now more usual for this mode of assessment (Moss, 1992), being neutral with respect to rectitude while suggesting that the test items should require some kind of active demonstration of the knowledge in question rather than a propositional account of it. Some action is required in a realistic setting, involving enactment of a skill or problem solving. Conventional pencil-and-paper tests evidently do not meet this requirement.

Performance assessment (PA) is closely related to the concept of situated cognition. Brown, Collins, and Duguid (1989) suggest that the only valid or powerful form of learning is what takes place in situated contexts. Schools, on the other hand, assume "a separation between knowing and doing, treating knowledge as an integral, self-sufficient substance, theoretically independent of the situations in which it is learned and used. The primary concern of schools often seems to be the transfer of this substance" (Brown et al., 1989, p. 32). Brown et al. do not consider the declarative knowledge taught in schools to be robust, so that in continuing to focus on the transfer of such knowledge, the school culture becomes inauthentic, providing students with ersatz activities.

Schools should instead provide students with a context and activities that lead to the construction of knowledge as it used; the notion of performance assessment thus appears integral to this position.

Although there is no doubt that context-based learning is much more powerful than learning disembedded content, there is a place in school for learning decontextualized content; indeed, it could be argued that this is what schools are mainly for (Biggs, 1992a). If knowing and doing were inseparable, there could be a problem in accounting for civilization; to know only through doing virtually requires each generation to reinvent the wheel. Perhaps, then, we are talking about two different things: the status of different kinds of knowledge and the most efficient means of learning any kind of knowledge.

Whatever means of teaching we adopt—inductive, problem-based, hands-on on the one hand versus expository on the other—there is much knowledge, school-delivered, that is legitimately propositional or declarative and that needs to be assessed as such, as well as being assessed in its real world applications. Thus although learning takes place most easily in situated contexts, schools exist precisely to help students learn decontextualized second-order symbol systems, and the declarative and procedural knowledge encoded by them. Learning these systems and contents is unlikely to occur in children's direct experience of the world, and it does not come easily, possibly for biological reasons (Biggs & Moore, 1993), but such knowledge needs to be learned, and the quality of that learning needs to be assessed.

In short, then, there is a place for decontextualized learning and assessment, just as there is for situated. So far the assessment of decontextualized learning has been in practice greatly overemphasized quite out of proportion to its place in the regular curriculum, but it would be equally unbalanced to entertain only situated learning and assessment contexts.

Let me now refer to a model that provides a useful framework for both modes of assessment: for charting the course of developmental assessment and for providing a generalizable language for discussing student performance in at least some performance assessment tasks (I prefer my own term *ecological*, but *authentic* has only just been deposed and further semantic quibbling is undesirable at this stage).

A Generalized Model of Qualitative Assessment

In the developmental model of assessment it is first necessary to chart the course of development of a concept or principle, so that the stages of development can be defined and the level at which a student is currently thinking determined. We thus need to describe what the learning will be like at any particular stage in its growth. This may be done on a topic-by-topic basis, as has been for some topics by Marton and his co-workers using the techniques of phenomenography (Marton, 1988; Ramsden, 1988), which involves probing interviews that usually reveal layers of understanding of the target concepts: a hierarchy of conceptions that can be used to form assessment targets.

A more general model would require us to define increasingly higher quality in terms of such as aspects as increasing complexity of structure, abstractness, economy or elegance of processing, originality of the response, and so on. Quality involves many different aspects according to the task in question, but one aspect that is common to most tasks is the structural com-

plexity of the learning outcome. Basically, there are two aspects to structural complexity: the amount of detail in the student's response (the quantitative aspect), and how well put together that detail is (the qualitative aspect). Both aspects are important and may be classified by the SOLO Taxonomy (Biggs & Collis, 1982; 1989).

SOLO, which stands for *S*tructure of the *O*bserved *L*earning *O*utcome, provides a systematic way of describing how a learner's performance grows in complexity when mastering many tasks, particularly the sort of tasks undertaken in school. A general sequence in the growth of the structural complexity of many concepts and skills is postulated, and that sequence may be used to guide the formulation of specific targets or the assessment of specific outcomes.

1. The task is not attacked appropriately; the student hasn't really understood the point and uses too simple a way of going about it (prestructural).
2. One (unistructural), then several (multistructural), aspects of the task are picked up and used, but are treated independently and additively. Assessment of this level is primarily quantitative.
3. These aspects then become integrated into a coherent whole (relational); this level is what is normally meant by an adequate understanding of the topic. Assessment of this level becomes qualitative if it is to pick up its nature.
4. The previous integrated whole may be conceptualized at a higher level of abstraction and generalized to a new topic or area; this too requires qualitative assessment.

These levels and the general structures incorporated in each provide a basis for assessing the quality of particular learning episodes. It should be noted that although levels 1 and 2 may be assessed quantitatively and additively, 3 and 4 require an interpretive or hermeneutic approach (Moss, 1994); it is not the components themselves that determine the quality of the outcome, but the whole that their interaction and integration defines.

The SOLO taxonomy may be used in two kinds of assessment format:

1. *Assessing open-end outcomes* is a procedure that is straightforward and moderately well documented (Biggs & Collis, 1982). Its application to marking assignments in a standard letter-grade system is given in Biggs (1992b).

Although SOLO may not exactly provide a common currency across tasks, it does provide a common way of thinking about performance in quite different tasks, which is why it may be particularly useful in assessing, for example, students' portfolios. The justification of each performance the student provides in a portfolio, and the pattern formed by the selections as a whole, will exemplify a SOLO structure, which will say something about the way the student thinks about the course: as several unrelated tasks or performances, or as different key tasks that reveal an integrated way of conceptualizing the course, or even inventive applications or generalizations that go beyond the course itself.

2. *Assessing in an objective-type format*, the ordered-outcome format (Masters, 1987) is less well documented and needs some discussion here.

The ordered outcome format looks like a multiple-choice item without the choice: the subitems are ordered in a hierarchy of competence and all require a response, each indicating a particular level in the competence hierarchy.

Masters (1987) used a phenomenographic hierarchy in his prototype of the format (see above; Marton, 1988), but the disadvantage is that phenomenography is highly content specific and may yield any number of levels, which makes it complicated to use in practice. Using SOLO the following criteria emerge for subitems addressing the stem topic at each SOLO level:

1. *Unistructural*: Contains one obvious piece of information coming directly from the stem.
2. *Multistructural*: Requires using two or more discrete and separate pieces of information contained in the stem.
3. *Relational*: Uses two or more pieces of information each directly related to an integrated understanding of the information in the stem.
4. *Extended Abstract*: Requires use of an abstract general principle or hypothesis that can be derived from, or suggested by, the information in the stem.

To illustrate, an ordered-outcome mathematics test was given to several hundred Year 7 students in each of two Hong Kong schools (Biggs, Lam, Balla, & Ki, 1988). The content of the test is unremarkable in itself, but the responses to one item are revealing in the present context (see Figure 1).

The two schools perform similarly up to multistructural level, but they diverge sharply thereafter, 48% of School B students obtaining correct on the extended abstract subitem: 48% correct compared to 6% of School A; the differences between the students in Schools A and B, whatever their genesis might be, are reflected only in the most complex cognitive processes. A conventional test comprising an aggregate of mixed items scored correct or incorrect would be unlikely to pick up this qualitative difference in the students' mathematical thinking. Further, there is little reason why a teacher of grade 7 would normally think to test beyond a multistructural level. Here the ordered outcome format forces the teacher to think upward when designing the test, and hopefully likewise the student when responding.



Toothpicks are used to make the above patterns. Four are used to make one box, seven to make two boxes, etc.

		FORM 1	
		School A	School B
UNISTRUCTURAL	a. How many toothpicks are used to make three boxes ?	96%	99%
MULTISTRUCTURAL	b. How many more toothpicks are used to make 5 boxes than used to make 3 boxes ?	74%	76%
RELATIONAL	c. How many boxes can be made with 31 toothpicks ?	57%	70%
EXTENDED ABSTRACT	d. If I have made y boxes, how many toothpicks have I used ?	6%	48%

Figure 1. An ordered outcome mathematics item (Biggs et al., 1988).

In pointing out the two traditions in educational thinking, quantitative and qualitative, Cole (1990) points out that each has had a useful history and a continuing presence, but that as each appears incompatible the "public understanding of education is hurt by allowing these two unconnected conversations about educational achievement to continue separately" (p. 5), so that an integrating framework is necessary to allow each to persist where it is effective and appropriate. SOLO appears to provide such an integration: in the early learning of many topics, quantitative aggregation is an appropriate or convenient way of assessing, but for higher-order, applied, and critical thinking (relational and extended abstract) qualitative assessment is more appropriate. The important thing is that modes of assessment appropriate to lower levels do not preclude higher levels or suggest to students that they can meet requirements by substituting lower for higher levels, which is the usual problem with backwash.

Backwash: Problem or Solution?

I now return to the question of backwash: the notion that testing drives not only the curriculum, but teaching methods and students' approaches to learning, usually adversely (Crooks, 1988; Elton & Laurillard, 1979; Frederiksen, 1984; Frederiksen & Collins, 1989).

These observations on the effects of backwash have been largely of the traditional measurement model framework, or in CRT conceived quantitatively. In the last case, backwash is explicit and actively encouraged insofar as the test deliberately becomes the target for teaching, as in measurement-driven instruction (MDI) (Popham, 1987). An important value question is raised here, as there is evidence that the success of such a strategy depends on how students typically go about learning: those who typically focus on and reproduce detail (a surface approach to learning) like the strategy and do well, but those who adopt a more academic or deep approach, originally better than surface learners, become frustrated and do progressively worse (Lai & Biggs, 1994).

The problem here is simply that the target of learning defined in the quantitative framework is of a low cognitive level, covering only the first two levels in the SOLO taxonomy. Would targets of higher cognitive levels promote backwash beneficial to instruction and encourage students to adopt deeper approaches to learning? There is some evidence suggesting this to be so. Tang (1991), for example, showed that different modes of assessment, final (short-answer) examination and a single-topic assignment, elicited quite distinct assessment preparation strategies, the assignment generally producing higher cognitive level strategies based on wide reading, collaborative learning, and problem solving, but overriding the question of the actual mode of assessment was the student's perception of what was required for optimal results. Further, students needed to have the procedural knowledge necessary to enact those perceived requirements; extended writing was a novel task to many of Tang's students, and lack of the appropriate skills prevented them from realizing what they could see as required.

Higher cognitive-level targets need therefore to be perceived as requiring high cognitive-level preparation strategies that are in the student's repertoire. Ordered-outcome testing should send the message that what is important is to think in increasingly complex ways about a topic, not to obtain a certain

number of correct items. The gap between a student's best response and the highest response in the hierarchy tells both teacher and student what still remains to be learned. Wong (1994) set parallel forms of a grade 11 math test using the traditional format and quantitative scoring, and the ordered-outcome format; he then interviewed students while they solved the different item-types. The difference he describes is that between novices (solving correctly but algorithmically) and experts (solving economically and originally from first principles), yet it was the same student who was both novice and expert, the variable being the test item type. It is important that further such studies be carried out, as this would provide much needed empirical as opposed to rhetorical justification for qualitative approaches to assessment.

Dimensions of Assessment

Performance/authentic assessment, then, is but one aspect of the development from traditional testing to the current situation with respect to alternative modes of assessment. To put this in perspective, at least three dimensions are involved in this evolution:

1. The *function* of testing: is it to rank individuals along some assumed trait, as in Taylor's (1994) measurement model, or to refer an individual's performance to a standard?
2. The *nature* of what it is that is to be assessed: a unitized fragment of performance or the growth of understanding?
3. The *context* in which the test item is placed: is it embedded in a context isomorphic to that in which the knowledge is or will be used in everyday life, or is it abstract and decontextualized?

Table 1 puts these points together.

The qualitative-quantitative dimension establishes the nature of what is to be assessed and how it may be reported. The measurement and standards models may operate within the quantitative, yielding NRT and CRT as traditionally implemented, but only the standards model may operate within the qualitative dimension, because NRT requires unidimensionality, whereas assessing an outcome qualitatively is a hermeneutic not a dimensional or measurement-based process (Moss, 1994). The original authentic debate raised the issue of whether assessment might best be situated or decontextualized. Decontextualized assessment would include the usual method of pencil-and-paper testing, which may be construed within either a quantitative framework, as is traditionally has been the case, or a qualitative framework, as is the case with most SOLO testing to date and with ordered-outcome testing. We then need to distinguish between testing the student's developing understanding of

Table 1
Dimensions and Modes of Assessment

	Model	*Context (examples)*	
		Decontextualized	*Situated*
Quantitative	Measurement	1. NRT (MC test)	4. PA (NRT) (?)
	Standards	2. CRT (Mstry Lrn)	5. PA (CRT)
Qualitative	Standards	3. SOLO (Develptl)	6. PA (Ecolgl)

a concept, particularly but not essentially of declarative knowledge, and the student's ability to involve that knowledge in a task that has ecological validity with respect to the learning goals. The issue in the last case is not so much what kind of understanding students have of the content, but whether the taught content can empower decision making in a real context (Maguire, 1990; Masters & Hill, 1988; Wiggins, 1989).

We are, then, left with six cells indicating modes and contexts of assessment.

1. *Quantitative-Measurement-Decontextualized*: Traditional NRT, which now has a sophisticated technology and the overwhelming approval of the measurement establishment and of most administrators. Best used for selection and other individual or group comparative purposes, relatively curriculum-free. The backwash generated is almost certainly the most deleterious to teaching and learning of all six modes.
2. *Quantitative-Standards-Decontextualized*: 1970s style CRT, mastery learning. Linked closely to curriculum and instructional strategy. Best used for testing basic or core skills, behavioral skills. The quantitative framework, however, limits its application to lower cognitive levels. Backwash suits students preferring a surface approach to learning, but counterproductive for students preferring a deep approach (Lai & Biggs, 1994). A model that in many senses straddles 1 and 2 is Item Response Theory (IRT) or Rasch Model (Wright & Stone, 1979), which is quantitatively framed and uses a version of trait theory, but the scoring does not depend on how other students perform as in NRT, or on the items making up the test as in CRT. However IRT is controversial even within the quantitative framework, but is open to most of the same criticisms.
3. *Qualitative-Standards-Decontextualized*: The developmental mode of assessment where the interest is in the growth of skills and concepts in themselves rather than in their applications, using pencil-and-paper testing rather than situated performances. Best used for finding the levels of understanding of basic concepts that students have attained so far; these levels could (and should) be incorporated into the curriculum as targets for instruction. Alternative framework research (White, 1988) and phenomenography (Marton, 1988) use situations and hierarchies specific to given topics, whereas SOLO uses a more general framework. Ordered-outcome testing, whether based on SOLO or any other growth model, also falls into this category. Backwash from such testing is likely to be helpful as it encourages teacher and learner to think higher.
4. *Quantitative-Measurement-Situated*: PA in NRT mode. This would involve setting tasks the grading of which is competitive; perhaps an athletics meet is an example. However, the assumption of the measurement model, that individual scores can be related to a stable trait, is really not what PA is about. For all intents and purposes, then, this is an empty cell.
5. *Quantitative-Standards-Situated*: many PA tasks could be scored and graded quantitatively, for example, number of correct applications or behaviors, use of rating scales. Such quantitative assessments are useful for communication or for manipulation in combining grades or determining cut-offs. Although this may seem open to some of the objections already raised against quantitative assessment, it is much clearer to both teachers and

learners that the numbers are used here for logistic purposes and are not an implied statement about the quantitative nature of the performance. The form of understanding being created by the context is appropriate to the real-world use of the knowledge in question precisely because it is situated.

6. *Qualitative-Standards-Situated*: Ecological PA where the assessment task is situated and evaluated in a context close to the learning goals. Grading would be in terms of qualitative categories of mastery or competence. These might be quite task specific, or based on a more general framework such as SOLO. Best used for learnings that are meant to have direct real-world applications, which accounts for much school learning but not all. Backwash is likely to be helpful for learning as the testing situation should be designed to mimic the learning objectives, thus excluding cynicism and minimizing test-wiseness, short-cuts, or surface learning.

No doubt other modes of assessment could be generated using different or additional dimensions. Perhaps one that is missing here is self- versus other-assessment; that is a particularly interesting pivot in the *situated* column. Nevertheless, the six, or rather five, cells generated here show a disproportion in terms of frequency of use in relation to the instructional context in which they are most useful.

Part of this disproportion is due a context that is inimical to the newer, alternative modes of assessment. The measurement establishment has in effect imposed a set of assumptions about the nature of educational measurement that is simply at odds with what should go on in classrooms; administrators have, moreover, encouraged this domination because the assumptions and practices of traditional measurement serve political and utilitarian administrative ends rather than educational ones (Wilson, 1994). Nowhere could this be clearer than in the current pressures toward competence-based testing currently being experienced in many countries (Biggs, 1994). The quantitative-decontextualized framework of assessment, whether CRT or NRT, has simply been hard to resist, either because of direct pressure or more simply from inertia.

Nevertheless, as the rapidly increasing interest and acceptance of PA shows, it is clear that the time has come for change, and increasingly teachers are recognizing this, perhaps more so in Canada (Bachor & Anderson, 1994) than in other countries. A real practical difficulty in the way of wider acceptance is not that practitioners see no need for change, but that they lack a framework for interpreting results and incorporating them into established grading schemes; the reliability and validity of the new techniques are not evident, although many forms of assessment and in particular report writing provide a heavy workload (Bachor & Anderson, 1994). The cost benefits of PA are simply not seen to be favorable. All this suggests that much further work is necessary.

What needs to be done to make PA and qualitative assessment in general more acceptable?

Some Current Problems

Two main issues warrant discussion: determining conceptions of reliability and validity acceptable to practitioners, and adapting institutional structures to accept and make performance assessment workable in classroom contexts.

Conceptions of Validity and Reliability

As noted, the assumptions underlying PA and qualitative assessment in general are quite different from those underlying the measurement model. It is therefore inappropriate to apply the same tests and standards of reliability and validity. Messick's classic (1989) review of validity set the tone for rethinking this issue, but this article is not particularly oriented to qualitative assessment. However, his central point, that validity is not a property of the test, but of the interpretations and uses to which test scores are put and their consequences, opened the way to considerable discussion and a good deal of consensus about reliability and validity specifically in relation to PA (Bachor, Anderson, Walsh, & Muir, 1994; Frederiksen & Collins, 1989; Haertel, 1991; Linn, Baker, & Dunbar, 1991; Messick, 1994; Moss, 1992, 1994; Shepard, 1993; Taylor, 1994; Wolf, Bixby, Glenn, & Gardner, 1991).

Perhaps the most fundamental insight is that validity is now seen as being grounded in the theory and knowledge base of learning and teaching, not in a science and technology of educational measurement, so that "in place of ranks, we will want to establish a developmentally ordered series of accomplishments" (Wolf et al., 1991, p. 63). The task of test construction, then, becomes rather different from the traditional model; one now requires a theory of learning unfolding longitudinally, the construct being tested well represented in the elicited test behaviors and explaining them and specific stepped targets for each curriculum topic. The question is no longer, Does a test measure what it is supposed to measure? but, Does it do what it is supposed to do? (Shepard, 1993). Thus assessment becomes contextualized, so that although accuracy, coverage, or representativeness of teaching goals in the test items and consistency are important, so too are fairness and adverse consequences, although whether these last questions belong in the domain of test validity or professional ethics in general is debatable, although their importance is not (Maguire, Hattie, & Haig, 1994).

Another fundamental shift from classic test theory is the role played by judgment and consensus both in establishing construct validity (an end to those days when a test was "valid for anything with which it correlates" [Shepard, 1993]), and in interpreting test scores. Thus, although many are concerned about the low reliability of portfolios, for example, Bateson (1994), it is important to recognize that the old additive model of compounding unreliability no longer applies. Specifically, a hermeneutic approach to drawing conclusions from test performances avoids this problem, as the aim is to arrive at a judgment by understanding the whole in light of the parts; it is not a case of judging single performances and then aggregating. An example is how a journal editor judges whether to accept or reject a manuscript on the basis of informed advice, even when rejecting a paper advocating a hermeneutic over an additive approach on the grounds that additivity is the accepted model (Moss, 1994).

Finally, both fidelity or ecological validity and fairness require that multiple routes to the same goal performance are allowed. As in a music competition, performers choose their own particular items, often using different instruments in order to show themselves at their best; analogically, then, students have the right to choose what to put in their portfolio (Moss, 1994). The final

decisions, whether summative or formative, are holistic and qualitative, requiring expert judgment.

For educators brought up on traditional test theory these considerations are likely to feel alien if not misleading. Undoubtedly Bateson (1994) is correct in saying "University preservice training has generally failed dismally in preparing teachers for the testing, assessment, and evaluation tasks which they must undertake in their classrooms" (p. 239). However, there is now a growing theory, and its technology, that could inform preservice and inservice teacher education and could replace the unsystematic but "amazing 'moccasin telegraph'" (p. 240) that spreads new and good ideas among teachers.

Institutional and Systemic Issues

The commitment and competence of teachers is certainly important in impeding change, but equally if not more important is the way institutions run. An educational institution is a system, that is, a working whole made up of a set of component parts, each of which affects the other until the whole forms an equilibrium, that state of equilibrium becoming the system (von Bertalanffy, 1968). The prime example of a system is, of course, an ecological system, in which individual component organisms lie in a delicate state of symbiotic interdependence. However, systems theory can apply to almost any complex situation, including schools and classrooms (Yinger & Hendricks-Lee, 1993). In the classroom the system comprises teacher, students, curriculum, teaching method, methods of assessment, and learning and teaching outcomes, all in a state of interdependence (Biggs, 1993). Part of the fierce resistance to change in assessment schemes is that assessment is so powerful a component, its backwash affecting the preceding chain of teaching and learning. Unless all participants are willing to change in the directions required, resistance will be considerable. The British Columbian situation (Bachor & Anderson, 1994; Bateson, 1994) illustrates both the difficulty and the possibility of change.

Reid (1987) points to three important components in the institutional system itself: the rhetoric, or the official aims of teaching; the technology, which would make possible the realization of these aims; and the social system of the institutions, which determines what is allowable in the institution. The social system comprises: the requirements established on a collegial basis, mostly informal but often given formal weight in Faculty meetings; the formal requirements of bureaucracy; and the requirements of the student body, which may be formal or informal. It is probably the social system with its various collegial, accountability, and managerial agendas that exerts most pressure on the assessment system in use (Biggs, in press).

The greatest pressures will be for using quantitative assessment. Most institutions require combining summative assessments at some stage: subunit results to obtain course results, course results to obtain year results. This puts almost irresistible pressure on teachers to use quantitative marking schemes, because marks are easily added up and averaged and make discrimination between students extremely easy. Profiling or other qualitative schemes could be used, but usually are not in the event. Most teachers, therefore, mark quantitatively, with the sort of results discussed above (Lohman, 1993).

Pressures toward decontextualized testing also exist in terms of convenience, security, and tradition. Institutionally, administrators feel it essential

to have standardized procedures, timed testing conditions, and cheat-proof security in the interests of fairness and accreditation and in anticipation of possible law suits. The public, and students too, see a face validity in this. As for teachers, it is difficult enough breaking the habits of a lifetime to redesign situated assessment, let alone under the conditions prescribed by the bureaucracy.

Thus changing the assessment system means setting up a new equilibrium, perhaps requiring a new technology, almost certainly requiring a new deal to be struck at all levels in the existing social system of the institution: with colleagues, with the bureaucracy, and with students. It is not impossible, as the British Columbian experience shows, but it is likely to be difficult.

Conclusions

It is no exaggeration to say that the theory and practice of assessing learning are currently undergoing a major paradigmatic change. It is not a matter of CRT replacing NRT, but the intersection of a variety of movements that, interestingly, have come from the broader educational canvas, not from the testing establishment itself.

Certainly the notion of CRT is coming to the forefront, but in connection with other views about the qualitative nature of higher order learnings and the situated nature of learning. Few of these ideas are particularly new, but their interaction becomes paradigmatic, suggesting heavily revised views of reliability and validity and new formats of testing. The critical single notion underlying this is, amazingly enough, that educational considerations should drive testing, and not psychometric, bureaucratic, or political ones.

Assessment occupies a key place in determining quality learning outcomes, but assessment practices are part of a wider picture that includes but extends beyond the responsibility of any individual teacher. An institution is a holistic, interactive system, which for its own management has many procedures in place with their own functional use. However, these procedures often determine teaching and assessment practices, which in turn influence students' perceptions of what and how they will learn.

There are three main factors impeding change. The first and simplest is know-how; many teachers may simply not know how to improve their assessment techniques. Second, and more subtle, is the probability that they don't know that they don't know. If someone has a quantitative mindset, really believing that we should teach, learn, and assess by numbers, then they won't even see that there is a problem. Finally, teachers have the institutional social system to deal with.

Obviously the road to better teaching, learning, and assessment is a complicated one that is beyond the control of educational researchers themselves. However, the history of assessment has shown the unfortunately negative effect the outdated measurement establishment has had on classroom practice. At least we can do something about that, and systems being what they are, perhaps we can then strike different and healthier equilibria than currently exist.

Acknowledgment

I am indebted to John Kirby, Bob Wilson, and one anonymous reviewer for their valuable feedback and advice.

Note

This article is based on an invited address to the Canadian Association for Educational Psychology, presented at the CSSE conference, Calgary, June, 1994.

References

Archbald, D.A., & Newman, F.M. (1988). *Beyond standardized testing: Assessing authentic achievement in the secondary school*. Reston, VA: National Association of Secondary Principals.

Bachor, D.G., & Anderson, J.O. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education, 1*, 63-93.

Bachor, D.G., Anderson, J.O., Walsh, J., & Muir, W. (1994). Classroom assessment and the relationship to representativeness, accuracy, and consistency. *Alberta Journal of Educational Research, 40*, 247-262.

Bateson, D.J. (1994). Psychometric and philosophic problems in "authentic" assessment: performance tasks and portfolios. *Alberta Journal of Educational Research, 40*, 233-246.

Biggs, J.B. (1973). Study behaviour and performance in objective and essay formats. *Australian Journal of Education, 17*, 157-167.

Biggs, J.B. (1992a). Returning to school: Review and discussion. In A. Demetriou, M. Shayer, & A. Efklides (Eds.), *Neo-Piagetian theories of cognitive development* (pp. 277-294). London: Routledge and Kegan Paul.

Biggs, J.B. (1992b). A qualitative approach to grading students. *HERDSA News, 14*(3), 3-6.

Biggs, J.B. (1993). From theory to practice: A cognitive systems approach. *Higher Education Research and Development, 12*, 73-86.

Biggs, J.B. (1994). Learning outcomes: Competence or expertise? *Australian and New Zealand Journal of Vocational and Educational Research, 2*, 1-18.

Biggs, J.B. (in press). Assessing learning quality: Reconciling institutional, staff and educational demands. *Assessment and Evaluation in Higher Education*.

Biggs, J.B., & Collis, K.F. (1982). *Evaluating the quality of learning: The SOLO Taxonomy*. New York: Academic Press.

Biggs, J.B., & Collis, K.F. (1989). Towards a model of school-based curriculum development and assessment: Using the SOLO Taxonomy. *Australian Journal of Education, 33*, 149-161.

Biggs, J.B., Lam, R.Y.L., Balla, J.R., & Ki, W.W. (1988, November 26-27). *Assessing learning over the long term: The "Ordered Outcomes" model*. Paper presented to Annual Conference, Hong Kong Educational Research Association.

Biggs, J.B., & Moore, P.J. (1993). *The process of learning*. Sydney: Prentice-Hall Australia.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). *Handbook of formative and summative education of student learning*. New York: McGraw-Hill.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-41.

Cizek, G. (1993). Rethinking psychometricians' beliefs about learning. *Educational Researcher, 22*(4), 4-9.

Clark, J., Scarino, A., & Brownell, J. (1994). *Improving the quality of learning: A framework for target oriented curriculum renewal in Hong Kong*. Hong Kong: Honkong Bank Language Development Fund, Institute of Language in Education.

Cole, N.S. (1990). Conceptions of educational achievement. *Educational Researcher, 19*(3), 2-7.

Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research, 58*, 438-481.

Delandsheere, G., & Petrovsky, A.R. (1994). Capturing teachers' knowledge: Performance assessment. *Educational Researcher, 23*(5), 11-18.

Elton, L., & Laurillard, D. (1979). Trends in student learning. *Studies in Higher Education, 4*, 87-102.

Frederiksen, J.R. (1984). The real test bias: Influences of testing on teaching and learning. *American Psychologist, 39*, 193-202.

Frederiksen, J.R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32.

Haertel, E.H. (1991). New forms of teacher assessment. *Review of Research in Education, 17*, 3-29.

Hoffman, B. (1962). *The tyranny of testing*. New York: Collier.

Lai, P., & Biggs, J.B. (1994). Who benefits from mastery learning? *Contemporary Educational Psychology, 19, 13-23.*

Linn, R., Baker, C., & Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher, 20*(8), 15-21.

Lohman, D.F. (1993). Teaching and testing to develop fluid abilities. *Educational Researcher, 22*(7), 12-23.

Maguire, T.O. (1990, May 31-June 1). *Grounded authentic assessment and teacher education.* Paper presented to the Second Conference on Classroom Assessment, Vancouver.

Maguire, T.O., Hattie, J., & Haig, B. (1994). Construct validity and achievement assessment. *Alberta Journal of Educational Research, 40,* 109-126.

Marso, R.N., & Pigge, F.L. (1991). An analysis of teacher-made tests: Item-types, cognitive demands, and item construction errors. *Contemporary Educational Psychology, 16,* 279-286.

Marton, F. (1988). Describing and improving learning. In R.R. Schmeck (Ed.), *Learning strategies and learning styles.* New York: Plenum.

Marton, F., Dall'alba, G., & Beaty, E. (in press). Conceptions of learning. *International Journal of Educational Research.*

Masters, G. (1987). New views of student learning: Implications for educational measurement. *Research working paper 87.11.* University of Melbourne, Centre for the Study of Higher Education.

Masters, G.N., & Hill, P.W. (1988). Reforming the assessment of student achievement in the senior secondary school. *Australian Journal of Education, 32,* 274-286.

Messick, S.J. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Messick, S.J. (1994). The interlay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Moss, P.A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research, 62,* 229-258.

Moss, P.A. (1994). Can there be validity without reliability? *Educational Researcher, 23*(2), 5-12.

Newman, F.M., & Archbald, D.A. (1992). The nature of authentic academic achievement. In A.R. Tom (Ed.), *Toward a new science of educational testing and achievement.* Albany, NY: State University of New York Press.

Popham, W.J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan, 68,* 679-682.

Popham, W.J., & Husek, T.R. (1969). Implications of criterion-referenced measurement. *Journal of Educational Measurement, 6,* 1-9.

Ramsden, P. (Ed.). (1988). *Improving learning: New perspectives.* London: Kogan Page.

Reid, W.A. (1987). Institutions and practices: Professional education reports and the language of reform. *Educational Researcher, 16*(8), 10-15.

Shepard, L.A. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20*(6), 2-16.

Shepard, L.A. (1993). Evaluating test validity. *Review of Research in Education, 19,* 405-450.

Tang, K.C.C. (1991). *Effects of different assessment methods on tertiary students' approaches to studying.* Unpublished doctoral dissertation, University of Hong Kong.

Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large scale assessment reform. *American Educational Research Journal, 31,* 231-262.

Trigwell, K., Prosser, M., & Taylor, P. (1994). Qualitative differences in approaches to teaching first year university science. *Higher Education, 27,* 75-84.

Von Bertalanffy, H. (1968). *General systems theory.* New York: Braziller.

White, R.T. (1988). *Learning science.* Oxford, UK: Blackwell.

Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership, 46,* 41-47.

Wilson, R.J. (1994, June). *Back to basics: A revisionist model of classroom-based assessment.* Presidential address, Annual Meeting, Canadian Educational Researchers Association, CSSE, Calgary.

Wilson, R.J., & Kirby, J.R. (1994). Introduction: Special issue on cognition and assessment. *Alberta Journal of Educational Research, 40,* 105-108.

Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of research in Education, 17,* 31-74.

Wong, C.S. (1994). *Using a cognitive approach to assess achievement in secondary school mathematics.* Unpublished master's thesis, University of Hong Kong.

Wright, B.D., & Stone, M.H. (1979). *Best test design: Rasch measurement.* Chicago, IL: MESA Press.

Yinger, R.J. & Hendricks-Lee, M.S. (1993). An ecological conception of teaching. *Learning and Individual Differences, 5,* 269-282.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, March 1995, 18-35*

Debora Barnett-Foster
Ryerson Polytechnic University

and

Philip Nagy
Ontario Institute for Studies in Education

A Comparison of Undergraduate Test Response Strategies for Multiple-choice and Constructed-response Questions

In this study the response strategies employed by a group of undergraduate chemistry students to answer multiple-choice and stem-equivalent constructed-response test questions were examined. A combination of written (N=300) and oral (N=21) reports from the students were analyzed. The written data were analyzed for format-related differences in solution strategies and error patterns. The oral reports were scored for the presence of typical problem solving behaviors. Analysis of the written data indicated no significant differences in the types of solution strategies employed nor in the types of errors committed across test format. The oral data, however, revealed evidence of subtler differences in the types and frequencies of problem solving behaviors.

Cette étude examine les différentes stratégies utilisées par un groupe d'étudiant(e)s en chimie au niveau universitaire pour répondre à des questions à choix multiples et à des questions construites selon un modèle où il y a plusieurs réponses possibles à une question-souche. On analysa une combinaison des réponses écrites (N=300) et des réponses orales (N=21) obtenues des rapports des élèves. Les données écrites ont été analysées pour relever les différences reliées au format des stratégies des solutions et des tendances des erreurs trouvées. Les rapports oraux ont été évalués pour la présence des comportements typiques concernant la résolution de problèmes. L'analyse des données écrites indique qu'il n'y a aucune différence significative dans les sortes de stratégies utilisées ni dans les sortes d'erreurs produites par les élèves entre les deux différents genres de tests. Les données orales cependant ont révélé la présence de différences plus subtiles dans les sortes et les fréquences de comportements utilisés en résolution de problèmes.

Introduction

The question of test format equivalence has been and continues to be an issue of some controversy. Since the evolution of the multiple-choice test, educational researchers have debated the relative merits of this format versus formats that require the student to compose or construct an answer. Comparisons of test format have involved studies of reliability, validity, success rate, item difficulty, knowledge retention effects, expectancy effects, and strategies for preparing for tests. However, one of the primary issues of interest has been the

Debora Barnett-Foster is a professor of applied chemical and biological sciences. Her interests are evaluation in undergraduate science education and test format equivalence.
Philip Nagy is an associate professor of measurement and evaluation. His interests are assessment of complex learning outcomes and data analysis.

investigation of the equivalence of traits or abilities measured by the tests. Although this has been an active area of research, the findings have been inconclusive. Some reports suggest that the formats measure different traits (Ackermann & Smith, 1988; Birenbaum & Tatsuoka, 1987; Ward, Frederickson, & Carlson, 1980). Yet others have found no effect of test format on the traits measured (Bennett et al., 1990; Bennett, Rock, & Wang, 1991; van den Bergh, 1990; Ward, 1982).

Traub (1992) argues that format effects can be expected in situations where the answers to test items provided in multiple-choice questions can be "recognized" or arrived at by some process of comparison or elimination not available to examinees answering constructed-response items. On the other hand, he suggests that items that require manipulation of data or ideas are not subject to format effects, because presumably examinees would have to perform these manipulations regardless of item format. If this is true, then the knowledge domain and the task demand under evaluation may influence the outcome of trait-equivalence research on the two test formats and may help to explain some of the apparent contradictions in the literature.

In most studies of trait equivalence the effect of format has been determined through statistical analysis of performance scores based on written test answers. Relatively little work has been done to identify and characterize the traits measured by the two test formats. Where it has been undertaken it has generally been confined to one type of test format and has often been based solely on written test answers. In most studies that have identified test response strategies employed to answer multiple-choice questions, the primary source of empirical data has been derived from statistical analysis of keyed options and distracters. There have been only a few investigations in which the reasoning processes were examined using think-aloud interviews (Bloom & Broder, 1950; Connolly & Wantman, 1964; Farr, Pritchard, & Smitten, 1990; Kropp, 1956; Norris, 1992). In the case of the constructed-response test format, most conclusions about test response strategies have been derived from statistical analysis of test subscores and error patterns in written test answers.

The commonly held assumption is that the constructed-response format requires the student to use some form of production strategy whereas the multiple-choice format requires only that the student discriminate among alternatives. However, the empirical evidence to date about these types of strategies is limited. The elimination or compare-and-delete strategy is often considered to be a predominant response strategy in answering multiple-choice questions; yet it has rarely been documented (Farr et al., 1990; Snow, 1980). It has also been suggested that the provision of answer options in the multiple-choice question might alter or truncate the solution path that would otherwise be used if there were no answer options to guide the solver. The solver may even work backward from the answer options to develop a solution path in a manner similar to that described by Hayes (1981). Although there has been considerable speculation about the type of response strategies employed to answer multiple-choice and constructed-response questions, there is very little empirical evidence to either support or refute these beliefs.

This investigation was intended to explore and describe the kinds of response strategies that students typically employ in answering questions posed

in multiple-choice and stem-equivalent constructed-response formats, using written and think-aloud procedures designed to elicit evidence of the students' thinking processes. In order to control for the potentially confounding effects of domain and task demand variables, this investigation was restricted to a single type of task in a specific discipline. Chemistry was selected because little work on format effects has been conducted in the sciences and none in the field of chemistry. In this discipline, tasks that involved quantitative problem solving were selected as being most typical of questions posed to first-year chemistry students.

Method

Overview

A combination of written test answers and oral reports of the examinees about their answers were used to compare the types of response strategies employed across test format. The written test answers were categorized and analyzed for differences in solution strategies and error patterns across test format. The oral reports were coded and analyzed for the thought processes, particularly with regard to problem solving behaviors associated with written solution strategies.

Subjects

A class of 300 students in a first-year introductory chemistry course was selected for this study. All students in the course had at least one secondary school credit in chemistry. Students in the course were registered in one of three undergraduate professional degree programs: Applied Chemistry and Biology, Environmental Health, or Food and Nutrition. The investigators neither taught the course nor had any other instructional role in it.

Three weeks prior to the final examination the students were informed of the purpose and nature of the project and were invited to participate; 261 agreed to allow their test answers to be analyzed. From that group 51 students volunteered to participate in a posttest interview. Of that group 21 were selected to represent three ranges of academic achievement based on secondary school chemistry grades: 76-99% (n=6), 65-75% (n=9) and below 65% (n=3). In addition three students who had been admitted to the institution as mature students were also interviewed.

Instrument

The test instrument was the final examination of the course. As shown in Table 1, the test was prepared in two versions that were distributed to randomly divided halves of the class.

Each version consisted of 15 multiple-choice questions and five constructed-response questions. Some of the multiple-choice questions (designated MCQ) simply requested the examinee to answer by selecting one of the five multiple-choice options provided. Other multiple-choice questions (designated MCQWR) requested the examinee to select one of the five options and to provide an explanation for the selected answer. The constructed-response questions (designated CRQ) asked the examinee to compose a response. For reasons of examinee convenience, and in order to separate the parallel questions as much as possible, the actual order of item presentation does not correspond to the order of the numbered sets. In all cases the multiple-choice

Table 1
Test Construction Design

Question Set	Content Area	Cognitive Level	Question Types Test A	Test B
1 (4 items)	molarity	(RCA)	MCQ	MCQ
	buffers	(CCA)		
	stoichiometry	(CCA)		
	kinetics	(RCA)		
2 (3 items)	molarity	(RAA)	MCQ	MCQWR
	kinetics	(RRC)		
	thermodynamics	(RRC)		
3 (3 items)	half-life	(RCA)	MCQWR	MCQ
	thermodynamics	(CCC)		
	thermodynamics	(AAA)		
4 (5 items)	gas laws	(CAA)	MCQWR	CRQ
	kinetics	(AAA)		
	crystal structure	(AAA)		
	stoichiometry	(CAA)		
	buffers	(AAA)		
5 (5 items)	gas laws	(CAA)	CRQ	MCQWR
	kinetics	(AAA)		
	crystal structure	(AAA)		
	stoichiometry	(CAA)		
	buffers	(AAA)		

Note 1. MCQ, MCQWR and CRQ represent multiple-choice questions, multiple-choice questions with response and constructed-response questions respectively. In a single row, the question sets are identical except for format.
Note 2. The cognitive level assigned by the three judges for each question is provided in column 3 as a series of three letters: R for Fact Recall, C for Comprehension, and A for Application.

items (MCQ and MCQWR) were presented before the constructed-response items.

The first question set (set 1) consisted of four multiple-choice questions (MCQ) that were identical across the two test versions. Question set 1 was designed to reveal performance differences in the two groups.

Question sets 2 and 3 were intended to provide an opportunity to examine the effect of requesting written explanations for the multiple-choice answers on the answer selection process. In set 2 the students who wrote test version A answered three questions in an MCQ format whereas the students who wrote test version B answered the same three questions in an MCQWR format. This pattern was reversed for question set 3, using a new set of three questions.

Question sets 4 and 5 were designed to provide a comparison of response strategies of students to the same questions posed in multiple-choice (MCQWR) and stem-equivalent constructed-response (CRQ) format.

Design of the Test Questions

Each multiple-choice question consisted of one correct response and four incorrect distracters. The distracters were chosen from a set of the most common errors produced by students tested with an equivalent constructed-response question. The errors were derived from archived examination papers and from a pilot study. Stem and option statements were examined to eliminate grammatical inconsistencies, inequality of length, and word cuing. Options that may have generated ambiguity such as "all of the above" or "A and B" were not used. Placement of correct responses was distributed over the five option positions so as to minimize serial position placement effects. Numerical answer options were sequenced in order of increasing value. An equivalent set of short-answer questions was developed using identical question stems.

Because the cognitive demand of a test question may affect test response behavior, the questions in this study were categorized according to the cognitive demand required to answer them. Questions were classified as requiring *fact recall, comprehension, or application* (Bloom, 1956) independently by three chemistry professors.

Questions belonging to sets 1, 2, and 3 were classified as fact recall (R), comprehension (C), or application (A) questions. All the questions from sets 4 and 5 were classified as application questions.

Administration of the Final Examination

Test administration conditions and instructions were identical for both groups and were consistent with typical examination procedures. Both groups were tested simultaneously. For MCQWR questions students were instructed to select an answer from the five options and to explain in writing their reasons for their answer selection. Students were also informed that the provision of an explanation was required in order to receive credit for the correct response. Students were permitted to answer the questions in any sequence they wished and were informed that there was no penalty for guessing.

Interview Procedure

After the examination 21 students attended a think-aloud interview with one of the investigators. The majority of interviews were conducted in a seven-day period following the written test. However, conflicts occasioned by examination and holiday schedules forced four interviews to be conducted one month after the written test.

At the beginning of each interview the think-aloud method was explained to the student, using Hayes, Flower, Schriver, Stratman, and Carey (1987) as a basis. Following this explanation a warm-up exercise was used to acquaint the student with the think-aloud technique. After the warm-up exercise the student was asked to read a handout that described the task about to be undertaken.

The student was then provided with an unmarked copy of his or her answers to the final exam, along with examination materials (periodic table, data sheet, calculator). The student was directed to examine certain test questions and to report, without attempting to explain or edit, all thoughts that had passed through his or her mind during the original problem solving session. In the event that the student could not recall how the problem had been solved

during the examination period, he or she was then instructed to attempt the problem and to vocalize all thoughts during this exercise. The provision of the written test responses during the oral protocols minimized the incidence of failure to recall.

During the interview the investigator intervened as little as possible. If the student paused for 10 seconds or longer, the investigator encouraged the student to continue to vocalize all thoughts. Otherwise the interviewer remained silent.

The meeting was tape-recorded and the investigator took notes. The sessions typically lasted one hour. At the end of the meeting all notes made by both parties were collected and collated with the copy of the subject's test.

The selection of test questions for interviews was designed to include for each subject three questions from each of set 2/3 and three from set 4 and set 5. Question selection was systematically varied across interviews so as to collect data on all questions in sets 2, 3, 4, and 5 for both test versions.

Results and Analysis

Analysis of Written Tests

For the purposes of this investigation, test responses for MCQ, MCQWR and CRQ were scored dichotomously. Differences in mean score and response pattern were examined for all written responses. Responses to question sets 4 and 5 were also analyzed for similarities and differences in the solution process used to generate the answer. The responses were first sorted on the basis of whether sufficient information was provided to categorize the solution processes used. Then those responses that were sufficiently complete were analyzed by question-content area for patterns in solution strategies and errors. The patterns were contrasted across question format.

Reliability coefficients (KR 20) were similar for both test formats. For the 15 multiple-choice items, the KR 20 values with means and standard deviations for versions A and B were 0.70 (8.6, 2.9) and 0.65 (8.7, 2.8) respectively. For the five constructed-response questions, the KR 20 values (means and standard deviations) for versions A and B were 0.74 (2.6, 1.7) and 0.64 (2.9, 1.5) respectively. It should be noted that the results for a single topic and format were combined across both test versions for analysis.

Given the sequential nature of some of the data analysis, with each subsequent stage being dependent somewhat on the results of the preceding stage, the description of the analysis and the presentation and discussion of the results are presented for each stage in order. Results of the written test analysis are reported by question set.

Question set 1. A student's t-test revealed a significant difference between the mean total scores computed from the four questions of set 1 for the two groups ($\bar{X}_{1=\ 2.1}$, $\bar{X}_{2=\ 2.4}$, $p \leq 0.01$). Using the two sample binomial test of equal proportions, analysis of performance differences between the two groups on each question revealed a significant difference only in the case of the buffer solution question. Because the mean score difference was only 0.3 of the pooled standard deviation (0.94) and a significant difference was found on only one question, the scores for sets 2 to 5 were not adjusted.

Question sets 2 and 3. A t-test analysis of the mean total scores for question set 2 (MCQ for test A and MCQWR for test B, identical questions) revealed no

Table 2
Response Choice Data (percent) for MCQ versus MCQWR (question set 3)

Question	Answer Option A	B	C	D	E	N/A
Half-life						
MCQ	4	79*	5	7	4	1
MCQWR	7	77*	5	6	3	2
Thermodynamics						
MCQ	0	1	56*	18	24	1
MCQWR	6	2	55*	17	21	0
Thermodynamics						
MCQ	3	39	39	10*	5	4
MCQWR	3	45	29	8*	4	11

*correct response.
N/A: no answer provided.

significant difference in performance between the two groups ($\overline{X}_1$= 1.4, $\overline{X}_2$= 1.5, p=0.15). A similar analysis for question set 3 also showed no significant difference in the mean total scores between the two groups (X_1=2.0, X_2=2.0). Because analysis of question sets 2 and 3 produced remarkably similar results, only the percent responses for set 3 are reported here in Table 2.

It is interesting to note that although there appeared to be a significant difference in the performance scores of the two groups for question set 1, the scores for sets 2 and 3 were not significantly different across the two groups. In the light of the crossed-design nature of sets 2 and 3, this finding provided support for the decision not to adjust scores for sets 2 to 5 based on the difference determined for set 1.

Chi-square analysis was used to determine differences in the frequencies of incorrect options chosen in question sets 2 and 3. The intent was to examine the effect of requesting a student to provide a written justification or reason for selecting his or her response. In particular, did this format lead to a change in the distribution of incorrect options selected? Chi-square values were determined for each question, first, with the four incorrect options and the no-answer response and, second, with the four incorrect options only. Because the two sets of data (including and excluding the no-answer response) differed only marginally, we decided to report only the results for the patterns of incorrect options selected. The Chi-square values revealed no significant differences in the patterns of incorrect responses across format for the six questions in sets 2 and 3 (Glass & Hopkins, 1984).

Question sets 4 and 5. Question sets 4 and 5 contained equivalent sets of multiple-choice and constructed-response questions designed to reveal differences in the response patterns for multiple-choice versus constructed-response questions. A two way (test x format) analysis of variance revealed no significant interaction between test and format (p=0.141). Test means for versions A and B were 2.79 and 2.89 respectively and format means for MCQWR and CRQ question sets were 2.96 and 2.72 respectively.

The five content areas were examined individually in a number of ways for differences across the two test formats. These included differences in proportion correct, proportion providing any answer at all, proportion providing an incorrect answer that was from among the multiple-choice options, and in patterns of particular incorrect answers. The first three of these analyses involved combining the data from question sets 4 and 5, whereas the last involved separate analyses of each set.

First, differences in proportion correct were examined across the content areas by combining the data from sets 4 and 5. Table 3(A) shows the difficulty levels of the two formats across the five content areas. The results reveal a significant difference, favoring the multiple-choice condition, for only one content area, crystal structure ($p<0.05$).

A second examination of the data from sets 4 and 5 concerns the proportion of students who failed to provide any answer at all. Table 3(B) contains the proportions of students who did not provide an answer under the two conditions. For all content areas, fewer students provided no answer in the CRQ format; the difference was significant only for the buffers questions, which were also by far the most difficult of the questions (Table 3(A)).

The third examination, reported in Table 3(C), concerns the proportion of wrong answers given by students in CRQ conditions that were not offered as multiple-choice options. What these data reveal is the success of the multiple-

Table 3
Differences in Proportions, Questions Sets 4 and 5 Combined (+ favoring MCQWR)

	MCQWR	*CRQ*	Δ
(A) Differences in Proportion Correct			
Gas Laws	.65	.60	+.05
Kinetics	.71	.74	−.03
Crystal Structure	.63	.51	+.12[1]
Stoichiometry	.56	.54	+.02
Buffers	.43	.34	+.09
(B) Differences in Proportion Not Answering			
Gas Laws	.02	.04	−.02
Kinetics	.01	.02	−.01
Crystal Structure	.02	.09	−.07
Stoichiometry	.02	.08	−.06
Buffers	.07	.28	−.21[1]
(C) Differences in Proportion of Incorrect Answers Not Provided as M-C Options			
Gas Laws	.00	.65	−.65[1]
Kinetics	.00	.38	−.38[1]
Crystal Structure	.00	.69	−.69[1]
Stoichiometry	.00	.70	−.70[1]
Buffers	.00	.71	−.71[1]

[1] $p<0.05$.

choice distracters in capturing the variety of likely wrong responses. Keeping in mind that these distracters were carefully constructed using archived constructed-response tests as sources of potential distracters, Table 3(C) reveals that in only one content area, kinetics, was this particularly successful. However, because the kinetics questions had the highest proportion of correct answers and hence were the easiest of the content areas, one would expect there to be less variation in the incorrect responses. In the remaining four content areas a great variety of other wrong answers were given by the students.

The fourth analysis examined the patterns across the two formats of incorrect responses that were provided as multiple-choice distracters. This analysis had to be carried out on each question set separately and did not take into account the "other" wrong responses. For questions in both sets, the Chi-square goodness of fit yielded nonsignificant values in all cases except one, indicating a reasonably good fit between the patterns of MCQWR errors and the patterns of those same errors in CRQ answers. Because both data sets yielded similar patterns, only the data for set 4 are reported here. Table 4 provides response percent data for all answer options.

The written responses to question sets 4 and 5 were also analyzed for similarities and differences in the solution path used to generate the final answer. The intent of this analysis was to examine differences in the solution paths leading to correct answers for multiple-choice and equivalent con-

Table 4
Response Choice Data (percent) for MCQWR versus CRQ (question set 4)

	Answer Option						
Question	*A*	*B*	*C*	*D*	*E*	*N/A*	*Other*
Gas Laws							
MCQWR	13	62*	14	8	1	2	0
CRQ	4	62*	4	0	0	2	28
Kinetics							
MCQWR	74*	4	9	7	6	0	0
CRQ	72*	1	6	2	4	1	14
Crystal Structure							
MCQWR	16	8	5	63*	7	1	0
CRQ	5	4	2	54*	1	7	27
Stoichiometry							
MCQWR	7	5	5	16	65*	2	0
CRQ	0	0	0	4	67*	7	22
Buffers							
MCQWR	6	15	13	40*	18	8	0
CRQ	0	4	4	34*	1	30	27

*correct answer.
N/A: no answer provided.
Other: answers other than those provided in the options.

structed-response questions and to compare across question format the solution paths that generated incorrect answers. Analysis of solution paths was possible only for those answers that were accompanied by a complete solution. Therefore, the written responses were first sorted for complete versus incomplete responses and then for correct versus incorrect final answers. Then the grouped response data were analyzed for similarities and differences in the solution paths used. Because this analysis was content-dependent, the response data had to be analyzed question by question. Results of this analysis were consistent for all question-content areas. For the sake of brevity, only one of the five question-content analyses, that for the partial pressures question, is presented in Table 5. For responses that produced a correct final answer to this question (5(B)), there was a marked similarity across question format in the type of solution path and frequency with which it was employed. For responses that generated an incorrect answer (5(C)), the type and frequency of strategies used were also similar across question formats.

Table 5
Analysis of Solution Strategies used for Partial Pressures Question

	Proportion of Answers	
A. Solution Responses for All Answers	*MCQWR (n=260)*	*CRQ (n=259)*
no answer	0.01	0.03
incomplete, incorrect/no answer	0.07	0.01
incomplete, correct	0.03	0.01
incomplete, incorrect	0.26	0.36
complete, correct	0.63	0.59
B. Strategies for Complete, Correct Answers	*MCQWR (n=164)*	*CRQ (n=153)*
added moles	0.46	0.42
added partial pressures	0.46	0.56
added volumes	0.01	0.01
made math errors, still got correct answer	0.04	0.01
illogical solution path	0.03	0.00
C. Strategies for Complete, Incorrect Answers	*MCQWR (n=68)*	*CRQ (n=95)*
used K_p formula	0.01	0.02
used Combined Gas Law	0.00	0.08
added weights not moles	0.29	0.30
made incorrect assumption	0.09	0.09
used incorrect molecular weight	0.34	0.28
solved for one partial pressure only	0.09	0.07
made math error	0.12	0.16
illogical	0.06	0.00

Analysis of Oral Reports

Coding scheme for oral protocols. The coding scheme used in this study to analyze the oral reports evolved from pilot interviews and the work of Hayes (1981) and Nurrenbern (1979). The final coding scheme consisted of four categories, which are described in the following sections A to D. Each category was defined by a series of typical behaviors. Because the categories were not mutually exclusive, the coding method did allow for a single statement to be classified in more than one category.

A. Problem representation. Information-processing theories of problem solving emphasize the importance of the generation of a problem representation or problem space (Gick, 1986; Greeno, 1978a, 1978b; Hayes, 1981; Newell & Simon, 1972). When constructing a representation of the problem the solver attempts to understand the problem by extracting the given information and connecting it to existing knowledge so as to identify the gap that must be crossed. In this study types of behavior that were classified as problem representation included rereading the problem, drawing a diagram, and defining the problem goals.

B. Strategy development. Evidence of strategy development was defined as any explicit reference through word or action to strategies to solve a problem or to steps of a solution path or formula.

C. Dominant strategy. A dominant problem solving strategy was identified for each subject on each problem. The strategy was based on the major sequence of processes used by the subject during the problem solving event. Where more than one strategy was used, both strategies were identified and labeled according to the sequence in which they were employed. The categories of strategies were derived from the model used by Nurrenbern (1979) and were modified to meet the goals of this investigation.

D. Evaluation processes. Each protocol was scored for types of evaluation behavior used to verify or assess the problem solution.

Coding reliability. The reliability of the coding scheme was assessed for both reproducibility and stability. Reproducibility or intercoder reliability was determined by comparing the protocol scoring decisions of two independent judges (one of the investigators and a third party) on four oral protocols. The intercoder agreement was 87.8% (123 out of 140 decisions made from the four protocols). Stability over time was determined by comparing coding decisions over an interval of two months; agreement was found on 89.1% of the 175 decisions made.

Results of Protocol Analysis

A data base of coded statements per subject-protocol was created. Because it was difficult to quantify the amount of a particular behavior, only the presence or absence of a behavior was scored. The frequencies for each behavior in a problem solving category were summed across protocols for each question format and question-content area. Then the frequencies for each behavior were summed in each question format across the five question-content areas. The frequencies for each of the four behaviors are reported in the two rightmost columns of Table 6.

Although the sample was too small to permit statistical analysis, it is interesting to note the differences in problem solving behavior across question

Table 6
Analysis of Verbal Reports (*N*=29)

	Frequency (%) MCQWR	CRQ
A. Problem Representation		
rereads problem or parts of problem	69	72
restates problem in own words	21	24
draws a diagram	10	14
makes a list	21	38
uses mnemonic notation	7	10
defines goals of problem	35	31
B. Strategy Development		
refers to algorithm/formula	83	86
refers to steps of a solution	59	62
recalls related concept	14	14
recalls related problem	7	24
reasons inductively/deductively	14	14
exhibits propositional logic	10	0
uses proportional logic	31	14
separates parts of problem	45	45
defines no strategy	7	0
C. Dominant Strategy		
algorithmic	62	83
algorithmic-reasoning	24	17
random trial and error	17	0
systematic trial and error	0	0
working backward	0	0
D. Evaluation Processes		
routine check of manipulations	7	28
checks that solution satisfies conditions	0	14
checks by retracing steps	3	7
derives solution by another method	14	14
is the result reasonable?	28	21
compares with generally known results	0	0
expression of confidence in answer	17	28
expression of lack of confidence in answer	24	14

format. For example, in the category of problem representation (Table 6(A)), there appeared to be a difference in the frequency of list-making behavior. Students who answered a question in constructed-response format tended to make lists more frequently than did students answering the same question in MCQWR format. This difference was more evident in questions that contained a considerable volume of data (e.g., the gas laws and stoichiometry questions). Also striking about this first section of Table 6 is that the predominant method of representing a problem, regardless of format, was to simply reread it.

In the development of strategy (Table 6(B)), students who solved MCQWR questions tended to use propositional and proportional logic more frequently than did students who answered the same questions in constructed-response format. This difference was traced to the kinetics question where proportional and propositional logic were most appropriate.

With respect to the dominant strategy used (Table 6(C)), the results indicate that most students used an algorithmic approach regardless of the question format and that none of the students employed either a systematic trial-and-error or a working-backward approach to the problems. It is interesting to note that the incidence of random trial and error, although small, was completely restricted to questions in MCQWR format.

The frequencies of different types of evaluation processes were too small to justify conclusions about any particular type (Table 6(D)). It was of some interest that the most frequently employed evaluation strategy was routine checking of manipulations. In order to get a sense of the extent to which evaluation activities were employed for the two test formats, the data for the first six categories were collapsed. A comparison of the frequencies of the collapsed categories across test format indicated a higher frequency of evaluation behavior with the constructed-response format. Also, a greater frequency of expressions of confidence was associated with constructed-response answers. Conversely, expressions of lack of confidence occurred more frequently with multiple-choice (MCQWR) questions.

Discussion

The results indicated that there was no significant difference in the success rate of students on the two question (MCQWR versus CRQ) formats as measured by mean total performance scores. The provision of multiple-choice options did not appear to enhance the ability of students to obtain the correct answer. From the written and oral reports it was apparent that students generally proceeded through a full solution process to determine the answer regardless of question format.

Analysis of the patterns of incorrect responses presented in Table 3 did reveal some format-related differences. A wider variety of errors occurred in answers to constructed-response questions. This was, however, not surprising given that for the multiple-choice questions the students were limited to four incorrect options. In fact the acid test of a multiple-choice question is its ability to provide incorrect options that comprise the majority of errors that students are likely to make in that particular problem. In this investigation the incorrect multiple-choice options were developed based on a review of the most common errors produced by similar groups of students to the same question posed in a constructed-response format. Even using this method to design the multiple-choice questions, the extent of other-answer production in the equivalent constructed-response questions varied from 38% to 71% of the group who incorrectly answered those constructed-response questions.

The question that produced the highest proportion of other-answer responses was the crystal structure question. From the oral reports, it was evident that most students had no sense of what would constitute a reasonable answer to the question. In other questions students appeared to be guided to some extent by the size of the expected answer and would reattempt the

solution if they produced what they felt was an unreasonable answer. For this question few students communicated a sense of what was reasonable. It was also apparent from the written and oral reports that many students had little understanding of the conceptual basis for solving this problem. This is consistent with the observation by Nurrenbern (1979) that students often solve chemistry problems by forcing the given information into a previously memorized formula (or in this case a formula provided on the data sheet). It should be noted that this problem was classified as an application problem and was not considered by the judges to be more challenging than the other set 4 and 5 problems.

When the answer pattern comparison was restricted to only those incorrect answers provided in the multiple-choice questions, the results indicated little difference between the two formats (see Table 4). Students appeared to make particular errors with similar frequencies for both question formats.

One question that did reveal a format-related difference in incorrect answer pattern production was the stoichiometry question. Because this question required a multilevel solution it offered more opportunity for error. Although there was no evidence of students being guided by the multiple-choice options in this question, the data suggested that different responses were elicited by the two question formats.

When the written explanations were analyzed, there appeared to be a significant difference across question format in the proportions of students who provided either no explanations or incomplete ones and yet still produced a final answer, whether it was correct or incorrect. The frequency of this type of response was significantly higher in the case of the multiple-choice question, suggesting that the incidence of risk-taking was increased for the multiple-choice format. However, it is quite possible that, in the case of the MCQWR, the students simply failed to provide the entire solution path that they used to generate the final answer. For this investigation the MCQWR and the CRQ were both scored dichotomously. However, for the purposes of student grading the mark allocation was MCQWR: 3 marks and CRQ: 10 marks. This difference in mark allocation may have affected the quantity and quality of the solution strategies provided in the written responses and possibly even the type of strategies employed. Although this problem might have been corrected or at least minimized by increasing the value of the MCQWR answers, such an adjustment would have created an atypical testing situation.

When the written answers were analyzed for differences across format in the solution paths used, the following patterns were noted.

1. There was little differentiation across test format in the types and frequencies of solution strategies employed to solve problems. Students appeared to use certain solution processes for certain types of questions with similar frequencies across test format. Most of the strategies were algorithmic in nature and followed standard application procedures as taught in class. Although evidence of strategies other than algorithmic may be difficult to discern from written work alone, the subsequent oral data indicated a high degree of consistency between the written explanations provided by the student during the exam and the oral explanations of the thought processes associated with answer production.

2. There was more differentiation across question format in the types and frequencies of errors committed during the solution process. The variety of errors committed in constructed-response answers was greater than that committed in the multiple-choice answers. Nevertheless, the frequency with which certain types of errors were made was in general similar for both formats.
3. There was limited evidence of increased use of solution truncation in the case of one multiple-choice question (the kinetics question). A number of students used the rate law equation to determine the order with respect to one reactant and then made an assumption about the value of the order with respect to the other reactant. The incidence of this response behavior was higher for the multiple-choice format than for the constructed-response format. However, the majority of students who solved the multiple-choice version of this question worked through the entire solution process in a similar fashion to those who solved the constructed-response version.

Interpretation of the oral reports revealed that the frequency of elements of problem solving behavior across test format was remarkably similar (see Table 6). Most students, regardless of format, engaged in some form of problem representation followed by an episode of strategy development. The extent and type of problem representation behavior was similar for both test formats. Only list-making appeared more frequently in the solution of constructed-response questions. Because many cognitive psychologists believe that problem representation is a critical aspect of problem solving and one that differentiates novices from experts, it is encouraging for multiple-choice proponents to note that for questions of this cognitive level and for this domain problem representation behavior was not differentiated across test format.

The extent and type of strategy development behavior in which students engaged did not differ dramatically across format. The most common type of strategy employed was an algorithmic one. It was surprising to note that the frequency of use of propositional and proportional logic was slightly greater in the case of the multiple-choice format. Logical reasoning strategies such as these are generally considered to belong to the higher-order thinking skills. Therefore, it may be somewhat surprising that these reasoning strategies appeared more frequently for multiple-choice questions, because it is often alleged that students tend to rely on lower-level thinking processes to answer multiple-choice questions. However, the sample size used in this study is too small to support a definitive conclusion.

There was no evidence of the working-backward strategy for either test format. Using this strategy the solver first identifies the projected goal and then works backward from that goal to develop a strategy. It is thought that working backward is easier with a multiple-choice question because the projected outcome is contained in a limited set of options. However, the findings of this study indicated no incidence of its use. This evidence is encouraging news for multiple-choice proponents. If we assume that working backward is an undesirable problem solving strategy, then its employment to answer test questions should be discouraged. Nevertheless, a much larger sample would be required to address this issue definitively.

With respect to the extent of evaluation processes employed during the test answer generation, the data indicated that evaluation processes were used more frequently to answer constructed-response questions. This may have been attributable to the greater mark allocation for CRQ. On the other hand, it may have been that the provision of answer options in the MCQ (and MCQWR) leads the solver to believe that an evaluation step is unnecessary. If this had been the case, one might have expected a higher incidence of expression of confidence associated with the multiple answers as opposed to the constructed-response answers. Yet the results showed the reverse in terms of frequency of expressions of confidence or lack thereof. Once again, the sample size is not large enough to provide a definite answer.

It is acknowledged that the time lapse between the written tests and the oral reports may have affected the quality and accuracy of the oral data. Ericsson and Simon (1980) indicate that oral reporting about a task becomes more variable as the time between the task and the reporting event increases. In this study the students were encouraged to simply report all thoughts that had transpired during the test task, and from the protocols it is clear that some students were more successful at doing so than others. In fact, what was vocalized undoubtedly involved some reporting and some reconstructing. The extent of new constructions was minimized by providing students with copies of their written test responses. Furthermore, a comparison of the reports generated during the week following the examination with those produced one month after the exam revealed no difference in level of detail or consistency with the written test answers.

Conclusions

This investigation has revealed that for questions that require some idea and data manipulation the solution processes employed to answer these questions in the two formats are remarkably similar for first-year university students in science and science-related programs of study. The findings support the recent hypothesis by Traub (1992) that questions that pose scientific or mathematical problems appear to be impervious to format effects. Yet these findings also indicate that at a finer level of examination there appear to be differences in elements of the problem solving behavior associated with the two test formats.

Given the pervasiveness of the written test as an evaluation instrument and the extent to which educators depend on the instrument for myriad decisions, the findings of this study have significant implications for educators and students as well as for researchers.

Educators may find the results of this study helpful in selecting student evaluation instruments. The findings suggest that for first-year university students pursuing science-related programs of study, response strategies for the two most common test formats are remarkably similar for the types of questions used in this study. It should be noted that these questions all involved multistep calculations and in that sense forced the students to proceed through the same set of steps to achieve the correct answer. For these kinds of questions the findings indicate that both test formats may be equally appropriate for assessing problem solving ability for this type and level of student. Educators should be cautioned, however, not to extend this interpretation to questions that test strictly fact recall or other types of reasoning strategies. Indeed, re-

search to date suggests that, for these kinds of tasks, format does make a difference (Traub, 1992). Future studies that incorporate questions dealing with fact recall and other types of tasks as well as those that examine other student samples could enhance the understanding of test response behavior.

Educators may also benefit from the insight provided by this study into the types of solution strategies employed to answer certain kinds of questions and the patterns and frequencies of errors committed. The high incidence of use of algorithmic strategies coupled with the low incidence of use of proportional and propositional reasoning strategies found in this sample of first-year university chemistry students should give chemistry educators cause for concern. It is a sobering reminder of the chasm that exists between problem solving success in a test situation and deeper conceptual understanding.

The results also have potential implications for students of chemistry. It has been suggested that students study differently depending on the type of test they expect (Loftus, 1971; Traub & MacRury, 1990; Treversky, 1973). However, according to the results of this investigation, for the kinds of test questions considered here students might be better advised to study in the same manner regardless of the test format expected, because it appears that they tend to use the same strategies to answer questions in both formats.

Finally, this investigation contributes to a better understanding of the effect of test format in the evaluation of specific traits. It has only recently been recognized that test format effects are domain- and task-specific, and investigators have now begun to define these effects for particular disciplines and types of tasks. The present finding of minimal format differences for chemistry undergraduates and for this task type provides an additional piece in the increasingly complex puzzle of trait equivalence of test formats.

References

Ackerman, T.A., & Smith, P.L. (1988). A comparison of the information provided by essay, multiple choice, and free response writing tests. *Applied Psychological Measurement, 12,* 117-128.

Bennett, R.E., Rock, D.A., & Wang, M. (1991). Equivalence of free response and multiple choice items. *Journal of Educational Measurement, 28,* 77-92.

Bennett, R.E., Rock, D.A., Braun, H.I., Frye, D., Spohrer, J.C., & Soloway, E. (1990). The relationship of expert-scored constrained free response items to multiple-choice and open-ended items. *Applied Psychological Measurement, 14,* 151-162.

Birenbaum, M., & Tatsuoka, K.K. (1987). Open-ended versus multiple-choice response formats—It does make a difference for diagnostic purposes. *Applied Psychological Measurement, 11,* 387-395.

Bloom, B.S. (Ed.). (1956). *Taxonomy of educational objectives, Handbook 1: Cognitive domain.* New York: Davis McKay.

Bloom, B.S., & Broder, L.J. (1950). Problem solving processes of college students. *Supplementary Educational Monograph, 73.* Chicago, IL: University of Chicago Press.

Connolly, J.A., & Wantman, M.J. (1964). An exploration of oral reasoning skills in responding to objective test items. *Journal of Educational Measurement, 1,* 59-64.

Ericsson, K.A., & Simon, H.A. (1980). Verbal reports as data. *Psychological Review, 87,* 215-251.

Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement, 27,* 209-226.

Gick, M.L. (1986). Problem solving strategies. *Educational Psychologist, 21,* 99-120.

Glass, G.V., & Hopkins, K.D. (1984). *Statistical methods in educational psychology* (2nd ed.) Englewood Cliffs, NJ: Prentice-Hall.

Greeno, J. (1978a). A study of problem solving. In R. Glaser (Ed.), *Advances in instructional psychology* (vol. 1, pp. 13-75). Hillsdale, NJ: Erlbaum.

Greeno, J. (1978b). Understanding and procedural knowledge in mathematics instruction. *Educational Psychologist, 12,* 262-283.

Hayes, J.R. (1981). *The complete problem solver.* Philadephia, PA: Franklin Institute Press.

Hayes, J.R., Flower, L., Schriver, K., Stratman, J., & Carey, L. (1987). Cognitive processes in revision. In S. Rosenberg (Ed.), *Advances in applied psycholinguistics* (vol. 2, pp. 176-240). Cambridge, UK: Press Syndicate of the University of Cambridge.

Kropp, R.P. (1956). The relationship between process and correct item-response. *Journal of Educational Research, 49,* 385-388.

Loftus, G.R. (1971). Comparison of recognition and recall in a continuous memory task. *Journal of Experimental Psychology, 91,* 220-226.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice-Hall.

Norris, S.P. (1992). A demonstration of the use of verbal reports of thinking in multiple choice critical thinking test design. *Alberta Journal of Educational Research, 38,* 155-176.

Nurrenbern, S.C. (1979). Problem solving behaviors of concrete and formal operational high school chemistry students when solving chemistry problems requiring formal reasoning skills. *Dissertation Abstracts International, 40(4986A), 167.* (University Microfilm No. DEM 80-05926)

Snow, R.E. (1980). Aptitude, learning, and instruction. *Cognitive Process Analyses of Aptitude, 1,* 27-64.

Traub, R.E. (1992). On the equivalence of traits assessed by multiple choice and constructed response tests. In R. Bennett & W. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 29-44). Hillsdale, NJ: Erlbaum.

Traub, R.E., & MacRury, K. (1990). Multiple choice versus free response in the testing of scholastic achievement (Original in German). In K. Ingenkamp & R.S. Jäger (Eds.), *Tests und Trends 8: Jahrbuch der Pädagogischen Diagnostik* (pp. 128-159). Weinheim & Basel: Beltz Verlag.

Treversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology, 5,* 275-287.

van den Bergh, H. (1990). On the construct validity of multiple choice items for reading comprehension. *Applied Psychological Measurement, 14,* 1-12.

Ward, W.C. (1982). A comparison of free response and multiple choice forms of verbal aptitude tests. *Applied Psychological Measurement, 6,* 1-11.

Ward, W.C., Frederickson, N., & Carlson, S.B. (1980). Construct validity of free-response and machine-scorable forms of a test. *Journal of Educational Measurement, 17,* 11-29.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, March 1995, 36-62*

Jeremy Hull
Four Directions Consulting Group

Ron Phillips
Phillips and Associates

and

Eleoussa Polyzoi
University of Winnipeg

Indian Control and the Delivery of Special Education Services to Students in Band-operated Schools in Manitoba

The article addresses the need for special education programs for Indian students in Manitoba and relates this to a review of special education program delivery and supports in band-operated schools in the province. Although indicators of need suggest that Indian students may have greater and more complex special education needs than the general provincial population, until recently few Indian students living on reserves in Manitoba have been identified as having special needs. The low identification rate may be related to both lack of resources and mistrust within Indian schools and communities of special education assessment methods and programs. A review of special education programs in five band-operated schools shows that there are problems in the development, monitoring, and operation of special education programs in these schools. Although in theory the financial resources available for Indian special education programs compare favorably with those available in provincial schools, in practice Indian schools are functioning in isolation without the benefit of a regional or provincial system of specialist, planning, and monitoring supports. Five interrelated steps are proposed that could improve special education services in Manitoba's Indian schools.

Cet article traite du besoin d'avoir des programmes d'orthopédagogie pour les élèves autochtones du Manitoba et rattache se besoin à la révision de la présentation du programme d'orthopédagogie et des programmes d'appui dans les écoles gérées par des groupes amérindiens dans la province. Même si les indicateurs des besoins suggèrent que les élèves autochtones auraient des besoins orthopédagogiques plus nombreux et plus complexes que ceux de la population provinciale générale, ce n'est que récemment que les élèves autochtones habitant les réserves au Manitoba ont été identifiés comme ayant des besoins spéciaux. Le taux diminué de cette identification est possiblement relié à la pénurie de ressources orthopédagogiques et à une certaine méfiance qui règne dans les communautés amérindiennes et les écoles autochtones vis-à-vis les méthodes d'évaluation et des programmes orthopédagogiques. L'étude de programmes en orthopédagogie dans cinq écoles gérées par les autochtones

Jeremy Hull has worked as an education and research consultant in Saskatchewan and Manitoba since 1973 and is currently a partner in a Winnipeg consulting company.
Ron Phillips is a private consultant specializing in Native education. He recently completed his doctorate in special education at the University of Arizona.
Eleossa Polyzoi is an associate professor of education. She has been involved in several special education program evaluations and is currently evaluating special education policy and practice in the Czech Republic.

démontre qu'il y a des problèmes qui existent en ce qui concerne le développement, la progression, et l'opération de ces programmes orthopédagogiques dans ces écoles. Quoiqu'en théorie les ressources financières apparentes disponibles pour les programmes orthopédagogiques dans les écoles autochtones comparent bien avec les ressources disponibles dans les écoles provinciales, les écoles autochtones fonctionnent en réalité dans un isolement sans recours aux bénéfices d'un système de spécialistes régionaux ou provinciaux, à la planification, et sans appui. Cinq étapes entrelacées sont proposées qui pourraient améliorer les services en orthopédagogie dans les écoles autochtones du Manitoba.

Over the past several years increasing concern has been expressed over the adequacy of special education service delivery for students attending reserve schools. Because of the isolation of many reserve schools and the lack of local specialists and support services, Indian[1] students are less likely than other students to be provided with appropriate programs to meet their special needs. Studies to date have shown that Indian students have a higher incidence of age-grade deceleration and a higher rate of early withdrawal from school compared with other students. Special educational services are rarely available in the students' native language. Results of standardized achievement tests also suggest that Indian students' skills in mathematics and language arts are much lower on average than those of Canadian students generally, and the gap between Indian achievement levels and those of the general population becomes greater as the grade level increases. Socioeconomic conditions in reserve communities also have a negative impact on students. Because of their families' social and economic circumstances, Indian students are at greater risk of leaving school before reaching grade 12. Again, this problem is compounded for Indian children with special needs.

The purpose of this article is to review the need for special education programs for Indian students in Manitoba and the current state of special education program delivery in band-operated schools. First, an overview is provided of the way in which the concept of *Indian control of Indian education* has been defined and how special education has been viewed within that definition. Second, various aspects of the need for special education services are discussed, including enrollments, success indicators, and socioeconomic conditions. Third, the delivery of special education programs in five band-operated schools is described. This is followed by a description of the current organizational structures supporting the delivery of special education to band-operated schools. This discussion focuses on the system for providing funding and support services to local schools, using the provincial system in Manitoba for comparison. The article concludes with a number of recommendations for improving the delivery of special education in band schools.

Indian Control of Indian Education: An Overview

The concept of Indian control of Indian education derives from the spirit of Aboriginal self-determination and from the various treaties that were signed between Aboriginal groups and the Crown. In exchange for rights to Aboriginal lands, the treaties promised among other things to provide schools on reserves when requested to do so (Morris, 1862). However, since the 1970s Indian organizations and leaders have been critical of the manner in which education has been provided to Indian communities. Criticism has focused on

the poor success of Indian students and lack of participation of Indian people in the decision making structures affecting Indian schooling.

In 1972 the National Indian Brotherhood (NIB, 1972) developed a policy paper on *Indian Control of Indian Education*. This paper, which was subsequently accepted in principle by the Minister of Indian Affairs and Northern Development, became the basis for developments in Indian education over the next 20 years. It dealt with philosophy of education, jurisdiction, programs, teachers, and facilities. However, the issue of special education was not specifically addressed.

The NIB (1972) paper strongly advocated that band councils have control of decisions and that parents have a prominent role in decision making, but the paper did not argue that band-administered education is always preferable to education in provincial schools. In particular, there was concern that the federal government had been making agreements with provincial school authorities to provide for Indian education without the consent or participation of the bands concerned. It was proposed that each band decide whether to operate its education program directly or to enter into an agreement with a provincial school board.

The NIB (1972) paper emphasized that the schools should reflect Indian values and should incorporate Indian history and language in the curriculum. In addition, the paper argued for the provision of nursery and kindergarten programs in the schools, expansion of secondary, postsecondary, and adult education programs in Indian communities, and provision of drug and alcohol education and prevention programs. The paper also called for Indian controlled cultural education centers, the training of Indian teachers and counselors, employment of Indian paraprofessionals, and improvement of educational facilities.

Since the NIB (1972) paper was first produced, there has been a dramatic shift to locally administered reserve schools. In addition, nursery and kindergarten programs offered for four- and five-year-olds as part of the school program are now standard, many Indian teachers have been trained and are now teaching on reserves, and much greater participation rates of Indian students in postsecondary programs have been reported. For example, the number of Indian postsecondary students in Canada doubled between 1985 and 1991, reaching more than 21,000 in 1991-1992 (INAC, 1992a, p. 39). Many of these postsecondary students have become teachers and are now teaching in reserve schools. As of 1986, three quarters of those employed in education on reserves in Canada were registered Indians (McBride, Gagne, & Atwell, 1990, p. 5). Most Indian schools now offer some type of Native studies program and many provide for Aboriginal languages as a medium of instruction or as a subject area. Many bands have established elected or appointed education authorities that have responsibility for administering the band's educational programs. And it is now the policy of INAC that agreements with provincial schools for the provision of education to Indian students must always involve the affected bands (McKnight, 1988).

Despite these advances, some feel that Indian educational authorities are simply administering a non-Indian school system. For example, Koens (1989) argues that immediately after an Indian school comes under local control, there

is a burst of enthusiasm leading to a superficial increase in symbolic Indian cultural content such as crafts and dancing. However, when these changes fail to reduce chronic student absenteeism and other educational problems, the effort to increase Indian content in the school is typically allowed to evaporate, and the school program becomes a "pale reflection of the provincial curriculum" (p. 37).

In 1988 the Assembly of First Nations (AFN) published a major policy review of Indian education. This was an updated and much more detailed statement than the NIB (1972) paper, but the central themes were much the same. It emphasized preservation of languages and culture, instruction in Indian values, preparation for "total living," that is, development of skills and abilities to function in a variety of occupations and cultural contexts, and the preeminence of the authority of the local band council in making educational policy.

The review identified a number of problems with the use of standardized tests in First Nations schools, noting that achievement tests are widely felt to be culturally biased and that they fail to take account of differences among provinces in curriculum. Biased tests were viewed as being responsible for channeling Indian students into inappropriate remedial, occupational, and special education programs. Moreover, the review found that many band schools did not have the capability of undertaking diagnostic testing to screen for students with learning disabilities or to identify gifted students (AFN, 1988, p. 80).

On the other hand, the AFN review found that most First Nations communities surveyed were concerned about the lack of appropriate special education services and funding. The review found that "special programs with intensive remediation" are required for the teaching of basic skills, for speech and language development, for development of learning skills, and for the education of multihandicapped and developmentally delayed children (AFN, 1988, p. 88). Concerns expressed about special education included:

- classification of Indian students as having special needs primarily as a means of obtaining additional funding;
- lack of student assessment practices;
- failure to obtain permission of parents before placing students in modified programs;
- lack of local control over placement decisions;
- failure to address the needs of gifted students;
- failure to screen for health problems that may result in learning disabilities; and
- lack of accountability of provincial education systems to parents of First Nations students in special education programs.

Many of the issues that affected Indian education 20 years ago are not yet resolved. Although band councils and Indian education authorities now have much more authority in administering educational programs than they used to, these authorities have not fully assumed control over program planning or have not had the resources to implement the programs they desire.

The area of special education is a case in point. As is argued below, there is reason to think that special education needs in Aboriginal communities might

be greater than those in other communities. However, the AFN review makes it clear that there is a great deal of distrust of the area of special education in Aboriginal communities and schools. The fear is partly that Aboriginal children are being inaccurately labeled as having learning disabilities and are then shunted into classes that give them few opportunities to develop. Special education is also an educational specialization in which outside experts such as psychologists make decisions about the children. Local school staff and parents may feel intimidated and left out of the process.

The fundamental issue affecting the control and direction of Indian school systems, including the provision of special education services, is how control can be exercised locally when the funding comes from the federal government. Related to this is how the school administration and board are to be held accountable for their success in meeting educational needs and goals of the community. With funding provided by the federal government and the board elected by the community, the logic of accountability is broken. As Paquette (1986) notes, "The reality experienced by most aboriginal people, who are not ratepayers in a public system, is one of an educational system all of whose capital and operating revenues, and most of whose human resources as well, come from the 'outside'" (p. 21).

This general issue of control and accountability is more sharply drawn in the case of special education services. The somewhat specialized and technical nature of special education makes it a more difficult area for parents and school boards to come to grips with. For the same reason it may seem more difficult to incorporate Indian values into special education services. Despite the increase in the number of Indian teachers, few of these teachers have specialist training related to special education. The continuing distrust of the role and value of special education programs and assessment methods on the one hand, and the problem of meeting the educational needs of the students on the other, present Indian parents and educators with a dilemma: Should Indian schools adopt the non-Indian system that they may distrust? And if not, what new system of dealing with student's special educational needs will be used?

Underlying this dilemma is the question, what are the special educational needs of Indian students? The following section attempts to provide an overview of this question.

The Special Educational Needs of Indian Students

The information on Indian enrollments, success indicators, and socioeconomic characteristics given in this section is intended to provide the necessary background for understanding the need for special education programs for Indian students. The statistics given below describe Indian students under federal jurisdiction whose permanent residence is on reserve, even though they may be attending a provincial school and may be living away from home temporarily. These statistics do not include Indian students whose normal place of residence is off reserve.

Enrollment Trends

Between 1975 and 1991 the number of registered Indian students living on reserves in Canada increased by approximately 35%, from about 72,000 to 97,000 (INAC, 1992a, p. 43). Indian students are enrolled in three different

types of schools: Schools operated directly by the federal Department of Indian Affairs (federal schools), schools operated by local school boards falling under provincial jurisdiction (provincial schools), and schools operated by band governments or education authorities, including tribal councils (band schools).

In 1975 41% of Indian students in Canada attended federal schools, but by 1991 this proportion had fallen to 6%. Over the same period the proportion of Indian students attending provincial schools fell from 53% to 45%, and the proportion attending band schools increased from 4% to 47% (INAC, 1992a, p. 43). In effect, most of the Indian student population in federal schools has been transferred to band jurisdiction, as local bands assumed control of formerly federal schools. The proportion of Indian students attending provincial schools has fallen slightly, reflecting the tendency of band-controlled schools to increase the number of grades offered on reserve rather than sending students out to provincial schools.

An even greater shift of Indian enrollments into band-operated schools has taken place in Manitoba over the past 15 years (see Figure 1). Most of the federal schools in Manitoba were transferred to band control during the 1980s and 1990s. As of September 1993, of the 45 reserve school systems in Manitoba only seven were federally operated and all of these were slated to come under local control by 1994. As can be seen in Figure 1, as of 1992 band school enrollments were up to about 12,000 and federal school enrollments were down to about 1,500. Indian enrollments in band schools now make up more than two thirds of all Indian students residing on reserves in Manitoba.

The number of Manitoba Indian students attending provincial schools, has remained constant at about 4,000 students for the past 12 years. These students may be living on reserve and attending a nearby provincial school or they may be temporarily living off reserve in private homes or in residences in order to attend school. A small number of reserve students attend institutions designed to meet special needs, such as the Manitoba School for the Deaf in Winnipeg or institutions designed for those who have severe behavior problems, such as Knowles Centre, also in Winnipeg. About half of the Indian students attending

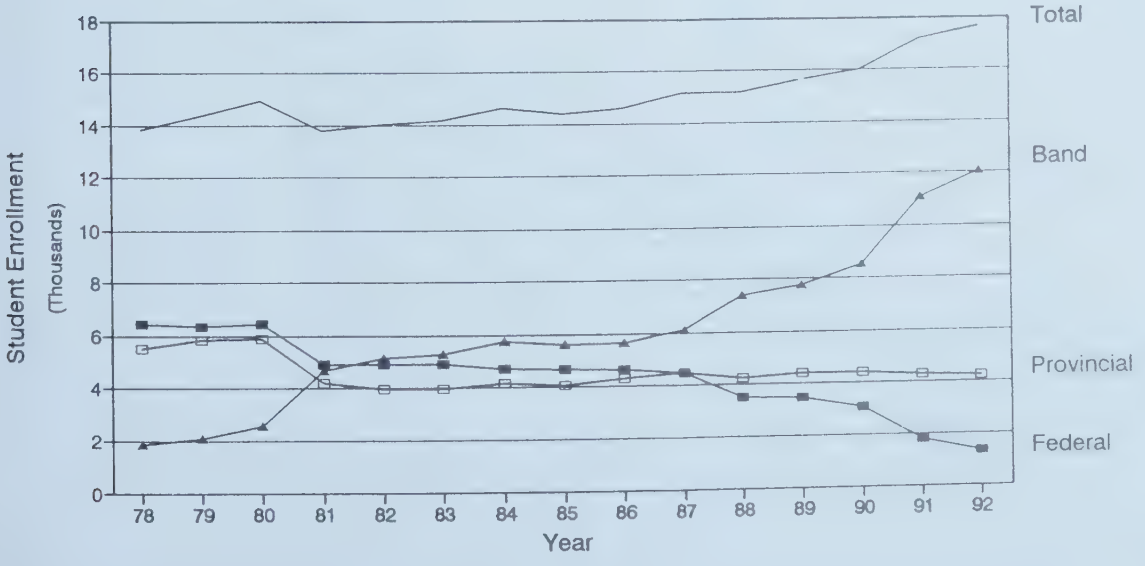


Figure 1. Enrollment of Indian students living on reserves by school type, Manitoba, 1978-1992 (source: unpublished Nominal Roll data).

provincial schools are enrolled in the Frontier School Division in the northern part of the province. More than one third of the Indian students attending provincial schools are enrolled at the high school level (unpublished 1992 Nominal Roll data provided by INAC, Manitoba Region).

Age-grade Deceleration

One indicator of the extent to which Indian students have special needs is the proportion of Indian students behind the expected grade level for their age group, referred to as age-grade deceleration.

Figure 2 compares the number of students of a given age who are one or more years behind the expected grade level for their age group in the three different school types during the 1986-1989 period. By age 12 more than half of the Indian students attending band and federal schools are decelerated by at least one year. The percentage increases dramatically in older students. This situation contrasts sharply with the degree of age-grade deceleration among the general student population of Manitoba, where it is not considered to be an issue.

The figure shows that there are slightly fewer students behind the expected grade in the provincial schools than in the band and federal schools. There are a number of possible explanations for this, such as differences in the social and educational expectations in different types of schools, differences between Indian students who choose to attend provincial schools as opposed to federal or band schools, and differences in schools' student promotion policies. To the best of our knowledge, no research has been done to date that provides a clear explanation for this difference.

At age 15 there is a noticeable increase in the proportion of students in federal schools who are behind the expected grade, and a decrease in the proportion of students who are behind in provincial schools. This may reflect selective transfers of more successful students from band or federal schools to

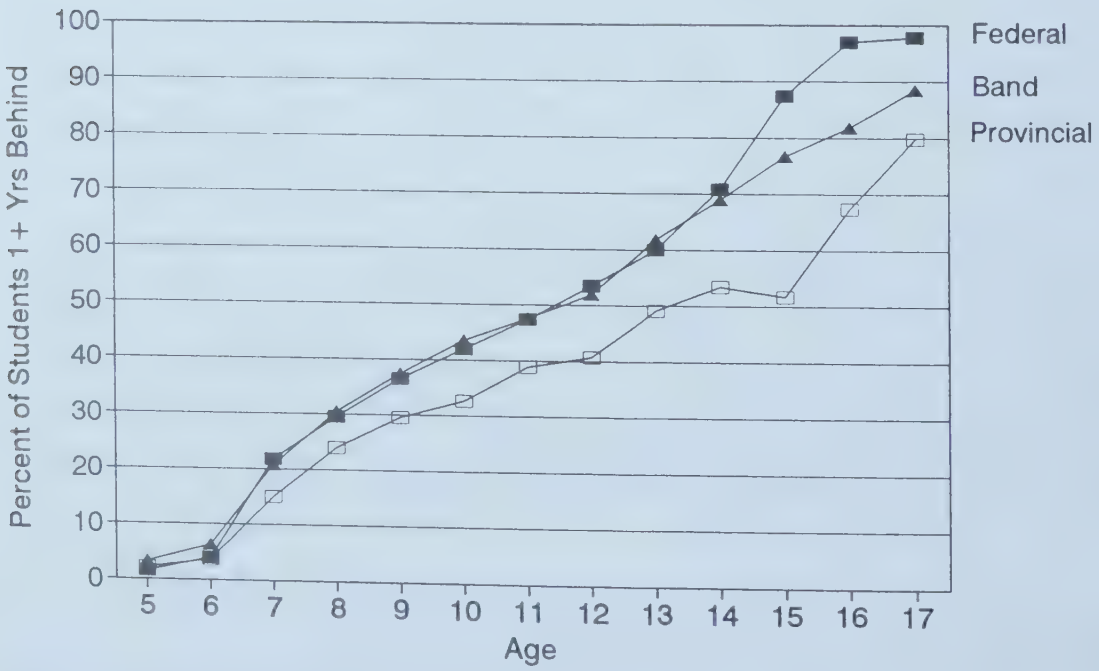


Figure 2. Manitoba Indian students behind grade by age and school type, 1986-1989 averages (source: unpublished Nominal Roll data).

provincial schools at about grade 9 or 10. It is therefore possible that those who remain in the reserve schools would be more likely to be behind the expected grade level and also more likely to withdraw prior to completing high school (personal communication, Director of Education of a rural band-operated school system, January 26, 1994).

Enrollment/Withdrawal Rates

A second possible indicator of special needs is the enrollment rate. This is defined as the percentage of students of a particular age group who are enrolled. It is, therefore, the converse of the dropout rate. An enrollment rate of 75%, for example, would suggest that 75% of the students of a given age are enrolled in school and that 25% have left school. The enrollment rates of Indian students are lower than those of the general provincial student population, but the gap between the two groups is much smaller than it has been in the past. Figure 3 compares enrollment rates among all on-reserve Indian students in Manitoba (regardless of the type of school they attend) with the enrollment rates of the general provincial student population.

As this figure illustrates, Indian enrollment rates decline rapidly after the age of 14 and are between 15% and 20% lower than provincial rates among 15-, 16- and 17-year-old students. For older students, Indian enrollment rates are higher than provincial averages, but this is partly because students in the provincial system tend to graduate from high school at an earlier age than do Indian students.

A number of studies of Indian student withdrawal rates were conducted in the 1970s. These studies reported average dropout rates across all grades of between 15% and 45% per year depending on the study (Stevens, 1982). Stevens reported that about 34% of Indian students in federal and provincial

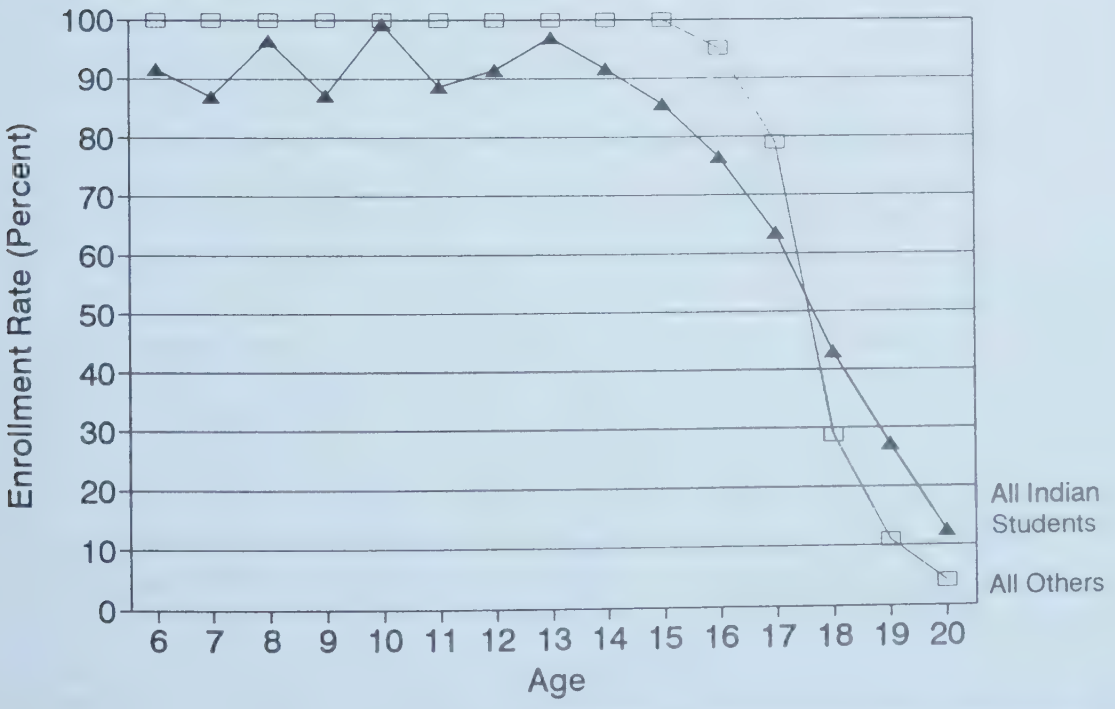


Figure 3. Manitoba K-12 enrollment rates, 1989-1990, Indian students and others, by age (source: Education in Canada, 1989-1990 [Statistics Canada #81-229]; Nominal Roll; Indian Register).

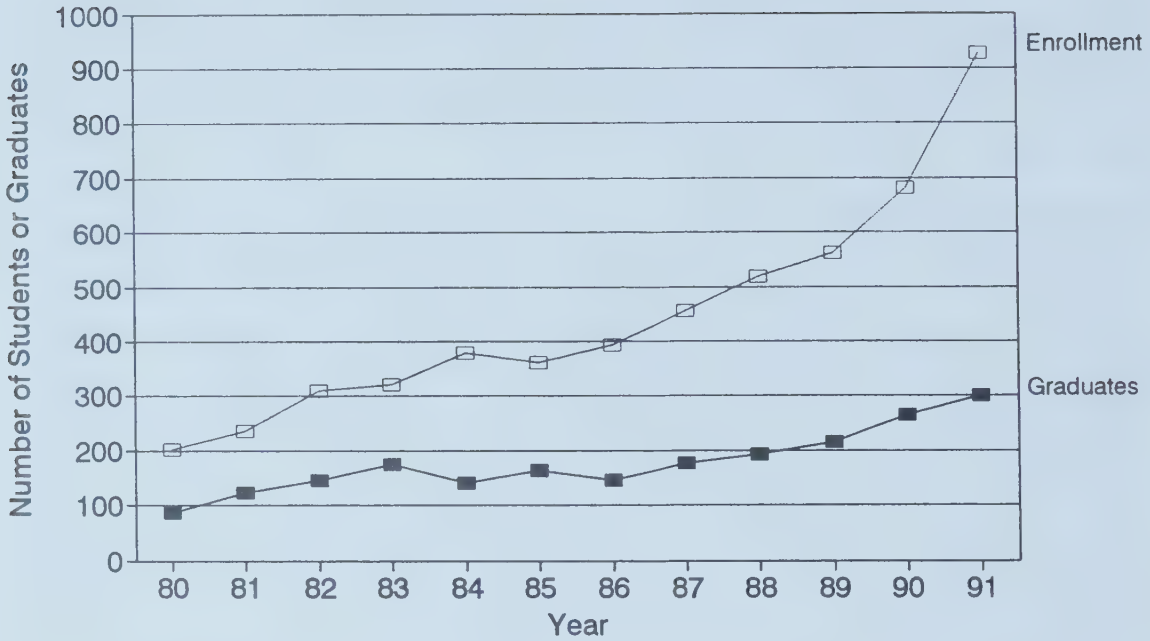


Figure 4. Indian grade 12 enrollments and graduates by school type and year, Manitoba 1980-1991 (source: unpublished data, Indian and Northern Affairs, Manitoba Regional Office).

schools in Manitoba dropped out each year during the 1977-1979 period. By comparison, the dropout rate of students in Winnipeg's inner city in 1978-1979 was 10% (p. 29).[2]

Graduation Rates

Another indicator of student success or levels of need is the graduation rate. As Figure 4 shows, both the number of Indian students enrolled in grade 12 and the number who graduate have increased over the past 12 years, but enrollments have increased more quickly than graduates. This means that the graduation rate (grade 12 graduates as a proportion of those enrolled in grade 12) has declined, from 43% in 1980 to 32% in 1991. The falling graduation rates may be a function of increased grade 12 enrollments. It may be that in earlier years only the most motivated and successful Indian students were enrolled in grade 12 and that the higher proportion of students enrolled in more recent years includes a larger proportion of students with somewhat lower motivation or academic skill levels. If true, this would be expected to lead to lower graduation rates.

Previous research found higher graduation rates among band-operated schools than among federal or provincial schools in Canada in 1980-1981 (Hull, 1990). However, in Manitoba for the period from 1980 through 1991 graduation rates in band-operated schools were lower than those in provincial or federal schools (INAC, Manitoba Region, Unpublished Nominal Roll Data, 1992b). It is possible that the lower graduation rates reflect greater success among band-operated schools in encouraging students to stay in school. This might have resulted from the combination of expansion of the number of grades offered locally (often associated with band-operated schools), selective transfers of better students to provincial schools, and the retention of students with weaker academic skills or less motivation in the band-operated schools.

Table 1
Percentages of Indian Students in Federal Schools and of All Students in Provincial Schools Who Met the Provincial Reading and Math Level Criteria Manitoba, 1980 and 1981

Grade	*Indian Students Attending Federal Schools*		*All Students Attending Provincial Schools*	
	Reading	*Math*	*Reading*	*Math*
3	44%	39%	80%	65%
6	35%	35%	62%	62%
9	42%	32%	62%	54%
12	48%	N/A[a]	61%	N/A[a]

[a]No test results reported for these groups.
(Source: Assembly of Manitoba Chiefs, 1984, pp. 85-86).

Achievement Test Results

Student success on achievement tests is another indicator of the level of need for special education services. Reading and mathematics skills of Indian students in federal schools in Manitoba were tested in 1980 and 1981 through the Manitoba Assessment Program. This program tests students in relation to provincial curriculum objectives, and was given to students in grades 3, 6, 9, and 12. As Table 1 shows, between 35% and 48% of the students in federal schools successfully met the reading level criteria, compared with 61% to 80% of Manitoba students in general. For both math scores and reading scores there was a large gap between the Indian students in federal schools, and the general provincial average (Assembly of Manitoba Chiefs, 1984, pp. 85-86).

No systematic study of Manitoba Indian students' performance on achievement tests since 1981 is available. However, a number of reserve school evaluations have been conducted over the past several years and many of these have reported achievement test results. Although various types of tests have been used in these evaluations, a number have been based on the Canadian Test of Basic Skills (CTBS). The CTBS produces a composite score that is a grade equivalent based on Canadian norms. Therefore, it has a built-in comparison with the achievement levels expected in Canadian schools. Although the CTBS has been criticized as inappropriate or culturally biased, it does at least provide a consistent reference point giving some indication of the success of students in gaining academic skills.

For the purposes of this article, evaluations of five schools in which CTBS results were reported at various grade levels were selected. The schools are not representative of all schools attended by Manitoba Indian students. They include only band or federal schools. Four are located in the northern administrative area of Manitoba. One is classified as urban, two are rural, one is remote, and one is isolated in terms of the INAC geographic classification system (INAC, 1987).

In most of these schools both the mathematics and language arts subtests of CTBS were administered. Table 2 presents these results in terms of the differences between the average grade equivalency scores of the CTBS tests and the

Table 2
CTBS Test Results Showing Average Years Behind Canadian Norms for Five Reserve Schools in Manitoba by Grade Level
Various Years from 1990 through 1993

Grade	Schools A	B	C[a]	D[b]	E	Average
1	.7	1.3	—	—	.5	.8
2	.8	—	—	—	.8	.8
3	1.1	—	.6	1.2	1.3	1.1
4	1.4	2.2	—	—	1.1	1.6
5	1.4	—	—	—	1.5	1.5
6	1.3	—	2.5	1.8	2.4	2.0
7	2.0	—	—	—	1.1	1.6
8	1.7	3.3	—	—	1.9	2.3
9	—	3.8	2.7	2.5	1.6	2.7

Note. All scores are composites of verbal and math scores except where otherwise noted.
[a]Reading comprehension subtest only.
[b]Math sub-tests only.
(Sources: Various unpublished school evaluation reports).

current grade level of the students. For example, if grade 3 students were tested at the end of November their current grade level would be 3.3 (three months into grade 3). If their average composite score was 2.5 the difference would be –.8 or eight 10ths of a year behind the norm in the skills tested.

The results shown in Table 2 indicate a trend toward greater differences in skill levels between Indian students in reserve schools and Canadian norms as the grade level increases. In other words, the longer the students are in school the farther behind they are falling. Although CTBS tests may not be sensitive to the cultural differences between Indian students living on reserves and other Canadians, they provide an indication of the success of students in obtaining skills that are valued in the job market and in Canadian society in general. Students moving from reserve schools into provincial school systems, postsecondary institutions, or the labor force are likely to face difficulties as a result of inadequate skill levels.

Socioeconomic Factors Affecting Schools

Socioeconomic conditions also provide an indication of the level of need for special educational services among the student population. The general pattern of greater educational attainment among those from higher socioeconomic status (SES) has been found to hold true for Indian students as well as others. An analysis of 1981 Census data found that as the SES index increased, the probability of 20-24-year-olds having completed high school also increased for both Indians and non-Indians (all others), and the gap between Indians and non-Indians became smaller (Hull, 1987a, p. 54). It was also found that in those Indian families with income below the poverty line, in those families where an Aboriginal language is spoken in the home, and among those living on-reserve, fewer children completed high school (Hull, 1987a, pp. 59-61). Low income

levels and high unemployment rates have been well documented among Manitoba's on-reserve Indian population (e.g., Hull, 1987b).

Special Education Enrollments in Reserve Schools

The above review of indicators suggests two things: First, there may be a higher incidence of special needs among students living on reserves than is found in other parts of Manitoba. Second, reserve students with special needs are attending school in a more difficult educational environment than other Manitoba students. Thus the difficulties they encounter because of their special educational needs are compounded by low levels of student achievement, high rates of withdrawal and age-grade deceleration, and difficult socioeconomic conditions.

Based on this reasoning we might expect a high proportion of students on reserves to be identified as having special needs. However, this has not been the case. Prior to 1986-1987, few Indian students were identified as having special needs, and the data available on special education students were limited and unreliable. Figures from the Nominal Roll during the 1978-1982 period showed that about 1.5 to 2.0% of Indian students were identified as special education students (Hull, 1987a). These figures seem unreasonably low since an estimated 10% of non-Indian students in Manitoba are believed to have special needs (Manitoba Education and Training, 1989).

In order to gain an independent estimate of the level of special needs in reserve schools, the Manitoba Indian Education Association (MIEA) conducted a survey in 1984 of 14 band-operated schools in seven Manitoba tribal councils to identify the need for special education services (Phillips & Cranwell, 1988). The survey involved personal interviews, primarily with teachers and principals, regarding the special needs of students enrolled in their schools. Results indicated that 969, or 31% of the 3,125 students surveyed, were suspected of having learning problems requiring specialized services. The survey was not intended to take the place of a thorough needs assessment, but it did provide an estimate of the total number of students who might potentially be in need of special education services.

In 1987 a report by the Island Lake Tribal Council on special education programs indicated that fewer than 1% of the students in their communities were identified as special education students on the Nominal Roll. Of this percentage, the majority tended to be classified by their medical condition. In addition, schools were often reluctant to refer students because parents were concerned over the negative effects of labelling, there were limited resources in the schools for properly assessing the child, and there was limited programming available to meet the students' needs once these needs were identified.

As recently as 1988-1989 it was found that 14 of the 32 band-operated schools in Manitoba and eight of the 13 federally operated schools were still not reporting any special needs students (Assembly of Manitoba Chiefs, 1991, pp. 82-83). These schools were located throughout Manitoba and ranged in size from seven to 747 students.

In 1986 reserve schools for the first time were able to apply for funds to provide services to their students identified as having special needs. INAC currently distinguishes between low- and high-cost special needs students. Low-cost special education funding for reserve schools is provided by INAC as

a per capita grant and is not dependent on the identification of specific individuals or needs. Indeed, low-cost funding is intended to provide services such as resource teachers, remedial programs, and assessment services that may serve a broad range of the student population. As of 1993-1994 the base allocation was $274 per full-time student per year (plus adjustments for reserve isolation and size factors).

Students can only be identified as high-cost after they are formally assessed by a psychologist or other specialist and only after the results of this assessment, accompanied by an Individualized Education Plan (IEP), are forwarded to INAC for approval. Once approval has been received, students are identified on the Nominal Roll registration form for the following year. Since 1991-1992, INAC has not required schools to formally retest students for whom a second consecutive year of funding is being requested (1993-1994 funding levels for high-cost students are shown in Table 5).

Since the new funding regulations came into effect there has been a noticeable increase in the number of formally identified special needs students attending reserve schools. Between 1985 and 1988 there was an increase of more than 200% in the number of Indian students identified as high-cost special needs in Manitoba (Working Margins Consulting Group, 1989, see Table 3). In 1988-1989 7.1% of the total student enrollment was identified as special needs. In Canada the percentage of the student population having special education needs has been variously reported in different studies as 9.07% (Canadian Council for Exceptional Children, 1989), 9.7% (Alberta Education, 1977), 12.5% (Roberts & Lazure, 1970) and 15.5% (Council of Ministers of Education, 1983). The comparable percentage for Manitoba is thought to be approximately 10% (Manitoba Education and Training, 1989). The relatively low proportion of identified special needs students may be attributed to the lack of special education resources, both human and financial, that are available for Indian students attending schools on reserve, which is in turn a reflection of the level of priority attached to special education services.

The above statistics, together with the review of indicators of risk among Indian students, provide an indication of problems in the identification of

Table 3
Growth in High Cost Indian Special Education Enrollments
Manitoba, 1985/86 to 1988/89

Year	*High Cost Student Enrollment*	*Percent of Total Student Enrollment*
1985/86	322	2.2
1986/87	471	3.2
1987/88	786	5.2
1988/89	1,075	7.1
Change (1985 to 1988)	+234%	

Note. Data are for Indian students whose permanent residence is on reserve. Students attending all school types (band, federal, and provincial) are included.
(Source: Working Margins Consulting Group, 1989, p. 48).

special needs students in Indian schools. However, it is necessary to look at the actual programs provided in specific schools in order to gauge the degree to which special needs are being met. The following section provides the results of such an examination for five band schools in Manitoba.

Evaluation of Special Education Programs in Five Band-operated Schools

In 1987 the Manitoba region of INAC established a funding program through which reserve schools were to be evaluated every five years (Mariash, 1989). This section summarizes the findings of five schools evaluations that took place between 1989 and 1992.[3] (In order to protect the anonymity of the schools involved, the following summary does not provide specific information concerning the numbers of teachers and students in individual schools.)

All but one of the schools were located in northern Manitoba. Three of the schools were in communities that did not have year-round road access. The enrollments of the schools evaluated ranged from about 80 to about 600 students. Three of the schools provided programs from kindergarten through grade 12, one school terminated at grade 8, and one at grade 9. Collectively, the enrollment of the five schools was 1,481 at the times of the surveys. Total teaching staff in the five schools, including teacher aides and principals, numbered 110, of whom a total of 41 were interviewed, including all teachers who were directly concerned with special education classes. Several other nonteaching staff were also interviewed in each school as noted below, such as home-school coordinators, nurses, band councillors, and others. A total of 115 special education students were identified in the five schools, including 50 in grades K-3, 38 in grades 4-6, 18 in grades 7 and 8, and 7 in grades 9-12. All available folders for these 115 students were examined as part of the evaluation process.

The following questions guided the evaluation of the special education programs at each school.

1. How are students referred for special education in the school? Are appropriate assessment and placement procedures used? Are these procedures being properly and consistently implemented?
2. How are Individual Educational Plans (IEPs) devised for special needs students, and are these adequately followed up?
3. What consultative or specialist resources and materials are used by the school in support of the classroom teachers? Are these resources available and adequate to meet the needs that have been identified?
4. To what extent have special education policies been developed concerning assessment, placement, informed consent, monitoring and follow-up of IEPs etc.?
5. Does the school have appropriate resource and curriculum materials to support the special education program?
6. To what extent are parents consulted with or involved in the special education process?

The special education evaluations included:

a. Informal on-site interviews with various teaching and administrative staff.

b. An examination of existing special education/resource folders of special needs students, including psychological reports and Individual Education Plans (IEPs).
c. Informal Interviews with various resource personnel in the Department of Indian Affairs, the tribal council responsible for placing selected special needs students in private homes in Winnipeg, and selected former staff of the school (in cases where key school staff had recently left the school at the time of the evaluation).
d. A tour of the school's facilities and selected observation of resource room activities.

The observations obtained during the evaluation of the five band-operated schools are summarized below. It should be noted that the five schools were experiencing similar problems with respect to special education. There was some variation in the quality of the Individual Educational Plans and in the nature of special needs programming that was provided, but there was a consistent lack of monitoring, specific guidelines for teachers, and follow-up of the recommendations. The comments below are generally true of all schools evaluated except where otherwise noted.

1. The process of identifying and assessing students with special needs may be described as follows. School staff in band-operated schools (resource teacher, classroom teacher, principal) target selected students for special education in May. Parental approval for comprehensive assessment is secured and a psychologist is contracted and flown into the community to assess the targeted students. The psychologist completes his or her assessment (she or he may also make a recommendation for a further speech and language assessment), writes a detailed report for each student and makes a recommendation regarding funding required from INAC. Based on the psychologist's report, the resource teacher develops an IEP for each student. A Student Specific Request Form, the student's Psychological Assessment Report and the IEP are forwarded to INAC as a request for special needs funding for the following academic year.
2. Reasons for referral typically provided in the students' folders included severe academic delay and stuttering (suspected Tourette's Syndrome), severe articulation difficulties, suspected cerebral palsy, and violent/aggressive behavior. Tests commonly administered by the psychologist included the Brigance Diagnostic Inventory of Basic Skills, the Torrence Test of Creative Thinking, the Peabody Picture Vocabulary Tests, the Wechsler Intelligence Scale for Children—Revised, the Bruininks-Oseretsky Test of Motor Proficiency, and the Key Math Diagnostic Arithmetic Test.
3. When the IEPs of the currently identified special needs students were examined, very little evidence of monitoring student progress was found. Individuals responsible for developing and carrying out student goals and objectives were not always identified on the IEP forms. In particular, the role of the classroom teacher in supporting the program developed by the special needs/resource teacher was not always delineated. Beginning and completion dates for specialized intervention related to each objective were also sometimes missing. As a result, IEPs seemed to be used primarily as a

means for obtaining funds from INAC rather than serving to assist the child in classroom intervention.

4. IEPs were sometimes too skeletal in form. In some cases, IEPs were photocopied and used for a number of students with similar needs. Although the intent may have been for teachers to expand and personalize these skeletal IEPs, evidence for this was not consistently present.
5. Although parental approval is required prior to assessment, this procedure was not always followed. For example, parents' signatures on approval forms were sometimes dated after assessments had taken place.
6. There were no specific instructional guidelines for teachers to follow in the students' Psychological Assessment Reports. The main recommendation in each of these reports was that the student be funded by INAC, but other recommendations often fell short of practical programming directives. The few recommendations that were provided were often too broad in scope. For example, for a student who displayed a severe language delay, it was recommended that "X be encouraged to copy, write and draw as much as possible to develop language skills and a language base." For a student who displayed very low self-esteem, it was recommended that "the teacher reward 'little successes' in the classroom." For a student who was achieving substantially below his grade level, it was recommended that he "be provided with individualized programming materials in the mainstreamed classroom and participate in small group activities, with the assistance of an aide."
7. Special needs programming for the identified high-cost students varied but primarily took the form of remedial instruction or alternative programming. Sometimes it was provided by the resource teacher on a pull-out basis; in other cases it was provided by the classroom teacher and aide. It was found that:
 a. Teachers were sometimes unclear about their role regarding programming.
 b. Teachers were not always involved in or informed about the students' IEPs.
 c. Withdrawal times for students to go to the resource room were inconsistent.
 d. Regular consultation meetings between the classroom and resource teachers were not always scheduled into their timetables. A collaborative/cooperative resource model was not always in place.
 e. Maintaining any specialized programming for a student was often difficult. The resource teacher often expected the classroom teacher to take more initiative; the classroom teacher expected the resource teacher to be continuously involved in the delivery of programming for the special needs student.
 f. In some schools there were limited resource materials, particularly for the upper grades. If special materials were purchased, they were not properly catalogued, making them difficult for other teachers to access.
8. Few students were identified as having special needs in the senior high school grades (grades 9-12, see Table 4). From 8.0% to 11.9% of students in the primary, intermediate and junior high grades were identified as special

Table 4
Percentage of Students Identified as Special Needs by Grade Level in Four Reserve Schools in Manitoba
Various Years Between 1989 and 1992

	School				
Level (Grades)	*A*	*B*	*C*	*D*	*Weighted Average*
Primary (K-3)	7.5	12.9	4.2	16.3	8.0
Intermediate (4-6)	5.6	12.3	16.9	5.0	11.9
Junior High (7-8)	3.1	6.1	13.3	0.0	8.5
Senior High (9-12)	1.6	0.0	2.4	—	2.2
Total	5.5	11.1	7.4	10.8	7.6
Total Identified Special Needs Students			104		
Total Enrollment			1,366		

(Sources: Data from various school evaluation studies completed between 1989 and 1992).

needs students, while only 2.2% of those in the senior high grades were special needs students in the four schools for which complete data were available. This pattern could reflect a greater dropout rate among special needs students, thus leaving few such students in the high school grades.

9. There was an evident absence of specialist support for the educational program. For example, if psychologists were flown in to an isolated community to conduct assessments, their primary role was assessment, not consultation on programming, teaching strategies or curricular resources. Many teachers expressed a need for more guidance in these areas.
10. This concern was compounded by a lack of training of resource teachers. Although all resource teachers were certificated teachers, few had special education certificates.

Organizational Support of Special Education Services in Band Schools

Apart from the funding supports for low- and high-cost special education students described above, INAC provides no additional funds for special education consultants, clinicians, or administration. INAC does provide one consultant and one psychologist who provide assistance to reserves throughout the province in establishing their programs. Bands typically provide a community-centered special education program consisting of resource rooms and special classes, such as remedial classes, life skills classes, and reentry classes for former dropouts, allowing students to stay in their home communities while attending school. However, the resource room and special class teachers and the regular classroom teachers operate without special education clinician and program support (see Figure 5). Lacking a system of coordinators or clinicians and other program supports, band-operated schools use private consultants to provide special education support services and student assessments.

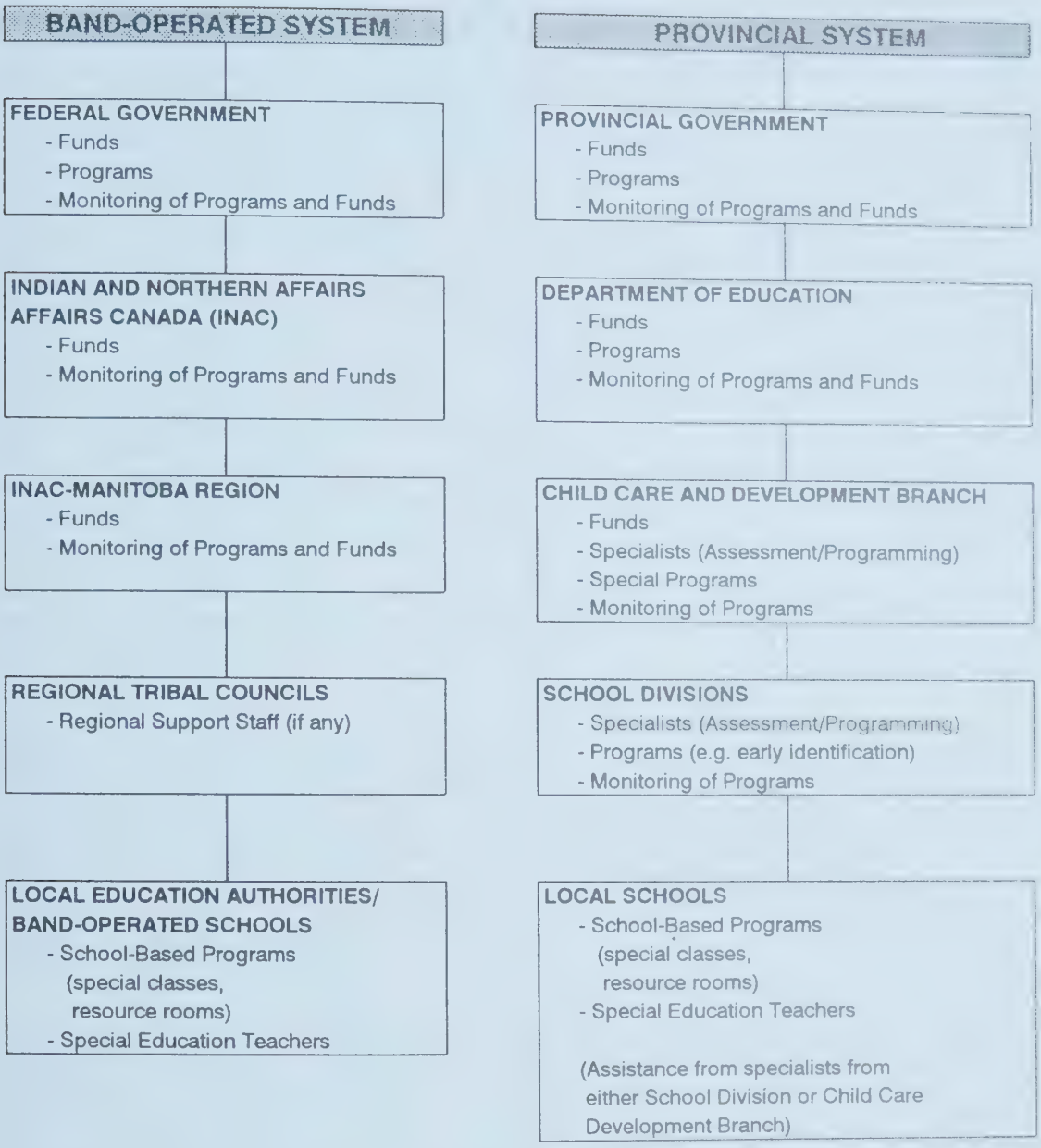


Figure 5. Comparison of band-operated and provincial special education support and monitoring systems.

Provincial (Manitoba) Special Education Policy

INAC support for special education, as with other aspects of the education program, is intended to meet provincial standards (Education and Research, 1985). Therefore, it may be useful to compare the INAC system of supports to the provincial system. Since 1967 the province of Manitoba has required school divisions or districts to provide educational programs for children with special needs. There is mandatory provincial legislation regarding special education (Canadian Council for Exceptional Children, 1992). The goal of special education in Manitoba is "to support children in developing the knowledge and skills they require to live meaningful, self-fulfilling lives with as much independence as possible in their communities" (Manitoba Education and Training, 1989, p. 3). This support is evidenced through a variety of special education programs and services to provincial school divisions/districts.

The province of Manitoba, through its department Manitoba Education and Training, provides special education funding to provincial school boards. The funding is provided on the basis of the total general student population as well as the number of individually identified special needs students (Canadian Council for Exceptional Children, 1992). Funding support for special education includes four components: coordinator and clinician support and Levels I, II, and III support (Manitoba Education and Training, 1992).

The coordinator and clinician support is based on the school division's total enrollment. This support is calculated as the "lesser of eligible enrollment divided by 700 adding one for any remainder or the number of qualified personnel employed, multiplied by $45,000" (Manitoba Education and Training, 1992, p. 9). In other words, divisions must spend this money on qualified personnel or they will lose it. Level I support is the lesser of the school division's enrollment divided by 180, multiplied by $45,000; or allowable expenditures. Level II support ($8,520 per student) is for individual students who have been identified as "severely multi-handicapped, severely psychotic or autistic or profoundly deaf" (Manitoba Education and Training, 1992, p. 9). Level III support ($18,960) is for individual students who have been identified as "profoundly multi-handicapped" (Manitoba Education and Training, 1992, p. 9).

Aside from special education support based on either total school enrollment or individually identified students, Manitoba Education and Training provides additional supports to provincial schools. These supports include direct consultation services for the visually and hearing impaired, vision/hearing screening, adaptive learning materials and special equipment, regionally based psychological and speech/language support services, the Diagnostic Centre (which provides assessment and programming), the Manitoba School for the Deaf, and professional development workshops.

In summary, the government of Manitoba, through Manitoba Education and Training, has developed a comprehensive system of special education

Table 5
Comparison of Provincial and Federal Funding Policies
(Prior to the 1994/95 Fiscal Year)

Type of Support	*Manitoba Education and Training*	*Indian and Northern Affairs (INAC)*
Base Funding	Enrollment/700 (rounded upward to nearest whole number) x $45,000	Enrollment x $274* (plus adjustments for isolation and size)
Level I Funding	Enrollment/180 x $45,000 (equivalent to $250/student)	$4,560/Identified Special Needs Student*
Level II Funding	$8,520/Identified Special Needs Student	$8,520/Identified Special Needs Student
Level III Funding	$18,960/Identified Special Needs Student	$18,960/Identified Special Needs Student

*As of the 1994/95 Fiscal Year, Base Funding and Level I Funding have been combined in a single grant of $400 per student. This is composed of $281 per student for Base Funding plus $119 per student for Level I Funding. The new funding policy also provides that no school will have its special education funding reduced as a result of the new formula.

Table 6
Special Education Low Cost and Level I Funding Available for Hypothetical Band Schools, Based on INAC Funding Policies
(Prior to the 1994/95 Fiscal Year)

Band-Operated School Enrollment	*Level I Students*		*Low Cost Grant*	*Level I Funding*	*Total Funding Available*
	Percent	*Number*			
200	5%	10	54,800	45,600	100,400
200	10%	20	54,800	91,200	146,000
600	5%	30	164,400	136,800	301,200
600	10%	60	164,400	273,600	438,000

services (see Figure 5). A wide range of funding, specialist, and program supports is available for students with special needs attending provincial schools. Schools, whether large or small, are able to obtain specialist and program assistance from either a school division or from Manitoba Education and Training.

Comparison of Federal and Provincial Special Education Systems

Support services provided by the provincial government for special education are different from those provided by the federal government. These differences are found both in the amounts of funding provided to schools and school divisions, and in the degree to which the respective governments exercise control over the way the money is spent. Table 5 provides a comparison of types of funding support provided in the two systems in Manitoba during the study period (1989-1992).[4]

It is difficult to compare the level of funding provided under the two systems, except by showing how they might work in practice with given enrollments. A typical smaller band school might have an enrollment of about 200 students, whereas a larger school might have an enrollment of about 600. From Table 4 it would be expected that each band school might identify 5% to 10% of their students as requiring special support, and most of these students would be assessed as requiring Level I support. Given this, the schools might be provided the amount of funding for special education (leaving aside Level II and III support) as shown in Table 6.

The comparable types of funding available through the provincial system to a given school are the clinician/coordinator support and Level I support. Provincial clinician funding is based on the collective enrollment of the school division. Therefore, the amount of support that might be considered as the share of the individual school would be calculated as a proportion of the division's total funding based on the proportion of the total enrollment attending the school. In Table 7 hypothetical enrollments for divisions as well as schools are shown, and these enrollments have been selected to show the possible range of situations that might occur.

Table 6 suggests that the funding available for special education programming, apart from Level II and III funding, in band schools with an enrollment of 200 might be from $100,400 to $146,000, depending on the number of iden-

Table 7
Special Education Funding Available for Hypothetical Provincial Schools Based on Manitoba Education Funding Policies

Provincial School Enrollment	Divisional Enrollment	Clinician Funding		Level I Funding	Total Funding[a]
		Division	School		
200	200	45,000	45,000	50,000	95,000
200	800	90,000	22,500	50,000	72,500
600	600	45,000	45,000	150,000	195,000
600	1,500	135,000	54,000	150,000	204,000

[a]School share of funding only.

tified Level I students. As shown in Table 7, for provincial schools the individual school's share of funding, apart from Level II and III funding, would be in the range of about $72,500 to $95,000, depending on the size of the school division in relation to the size of the school.

For a band school with an enrollment of 600, special education funding available might be from $301,200 to $438,000 (see Table 6), whereas for provincial schools of this size it might be from $195,000 to $204,000 (see Table 7). These figures do not take into account additional amounts provided to band schools because of small size or geographic isolation.

In spite of similar or higher funding levels for special education at band schools in comparison to provincial schools, in practice the overall system of supports for special education is much stronger in the provincial school system. Provincial schools all have access to special education consultants, clinicians, or coordinators through their school divisions. Provincial schools can also obtain more specialized assistance from regionally and provincially based consultants employed by Manitoba Education and Training. Some additional services are provided directly by the province, such as the Diagnostic Centre and professional development workshops. In short, the combination of school level, divisional, and provincial supports for special education results in a comprehensive system that provincial schools can access.

In contrast to the provincial system, INAC funding guidelines do not specifically provide for such services as administrative support, back-up specialist support, and early identification programs, nor does INAC provide a regional support system (see Table 8).

The lack of coordinators and clinicians in the band-operated system results in:

1. funds being spent on assessments by outside private consultants;
2. an absence of ongoing program consultation to special education and regular classroom teachers; and
3. lower numbers of identified special education students.

Given the funding policies reviewed above it appears that Indian education authorities have access to the financial resources required to provide a more comprehensive special education program, at least in theory. The question,

Table 8
Comparison of Special Education Services Provided to Students Attending Provincial and Band-Operated Schools in Manitoba

Special Education Service	*Provided in Provincial Schools?*	*Provided in Band-operated Schools?*
Remedial Education	Yes	Yes
Student Assessment	Yes	Yes
High Incidence Funding	Yes	Yes
Low Incidence Funding	Yes	Yes
Itinerant Teachers	Yes	No
Administrative Support	Yes	No
Specialists for Assessment and Programming	Yes	No
Early Identification Programs	Yes	No
Regionally-based Support	Yes	No
Provincially-based Support	Yes	No

(Source: Tom Walker, Director of Education, Swan Lake Education Authority, personal communication, Nov. 12, 1990).

then, is why is there evidently not an equivalent system in support of band schools? At least five factors seem to be involved.

Band schools are small and tend to work in isolation. Half of the schools located on reserves in Manitoba have enrollments of fewer than 200 students, and more than 70% have enrollments of fewer than 300 (Unpublished INAC Nominal Roll data, 1992). Moreover, these small schools are scattered throughout the rural and northern areas of the province. Although there is a system of tribal councils to which most Indian bands belong, the tribal councils have little responsibility in the area of elementary-secondary education. Only one tribal council in Manitoba has organized a school division (the Southeast Tribal Division for Schools) to provide central administrative and educational supports for a group of four reserve schools.

Local knowledge of special education, and local planning skills generally, are weak. As was found in the evaluation of five special education programs, there is often a need for professional development in the area of special education. Small schools may often lack the expertise to develop a special education program or even to organize systematic assessment of students. In order to both obtain and effectively utilize the special education funds available to them, band-operated schools would need to systematically undertake testing of students and would have to carefully allocate the per capita grant to either hire a specialist to do such testing or contract for such testing to be done. Once special needs students are identified and funding is received, this funding could be used to further enhance the special education program and services in the school. In effect, this is a two-stage funding process that involves delays and uncertainty as the school awaits funding approval by INAC. This process presents a barrier for some schools wishing to access funding.

Unlike provincial school divisions, band-run schools are not required to employ clinicians or to meet established program standards. As noted above, provincial

funding policy requires school divisions to employ qualified special education personnel in order to receive special education funds earmarked for the purpose. Bands, which are not subject to such a requirement, may use special education money for other purposes at the expense of the special education program. At present no system monitors whether this is happening.

Neither INAC, the various tribal councils, nor the chiefs have established any regional or province-wide education service organizations in support of band schools. Individual bands could allocate part of their education funding to establish regional support organizations that might be equivalent to the provincial system of supports. INAC could, as well, reserve part of the special education funds available for the support of such a regional organization. However, this has not happened, and it is not clear that bands or Indian education authorities would agree to such an arrangement as they would have to relinquish part of their local educational budgets.

Although funding for identified Level I students in band schools appears to be available, there may be a cap on this funding regardless of demonstrated need. Some school administrators believe that if they exceed the expected identification rate of 10%, they might not be able to access additional funds from INAC regardless of the demonstrated need.

Conclusions and Recommendations for Improving the Delivery of Special Education in Band-operated Schools

The foregoing discussion leads to four general conclusions. First, there is evidence that the need of Indian students for special education services may be greater and more complex than the needs of the general student population in Canada. Evidence of greater needs among Indian students takes the form of higher rates of age-grade deceleration, higher withdrawal rates, lower scores on achievement tests, and less favorable socioeconomic circumstances among Indian students in comparison with others. The greater educational and social needs of Indian students are complicated by language and cultural differences that may or may not be appropriately taken into account by the schools. There is little evidence that special education services in particular are provided in a way that has taken the social and cultural context of Indian students into account.

Second, Indian communities and schools are ambivalent in their attitudes toward special education programs. There is much concern about negative labeling and streaming of Indian students. Parents, educators, and Indian leaders are wary of remedial programs that tend to limit Indian students' opportunities and are concerned about the disproportionate number of Indian students who have been placed into such programs in provincial schools. They are also critical of standardized tests that are based on Canadian norms and are believed to be culturally biased. At the same time, those working in Indian schools are keenly aware of a variety of special needs among their students and wish to have the best, most suitable program to meet these needs. Moreover, identification of special needs students is a way of obtaining additional resources for the school.

Third, there are serious problems in the delivery of special education programs in Indian schools. The problems with the delivery of special education programs in the five schools evaluated were remarkably similar. These

problems involve the definition of the special education program, the roles and responsibilities of staff, the direction and monitoring of psychologists' work, the monitoring and followup of students, and the involvement of parents in the process. These problems may be related in part to the ambivalent attitude mentioned above, which tends to make special education programming a low priority in the school. In turn, the lack of development of special education programs means that special education programs in Indian schools are often based on outdated remedial program models, which may reinforce the community's negative view of the value of special education programs.

Fourth, there is no regional support system to assist local Indian schools as they attempt to identify special needs students and develop programs for them. Currently there is no system of monitoring the implementation of special education programs, apart from infrequent overall school evaluations. The lack of regional support or monitoring again undermines the local program and tends to reinforce negative perceptions of special education. It also tends to isolate each school system, and with many Indian schools having very small enrollments, there is little room in the budgets of these schools to hire specialists or address special needs.

In summary, local control of Indian education has not solved the problems of special education program delivery in reserve schools and has only recently begun to address them. The energy and enthusiasm that sometimes accompanies the movement to local control can neither be expected to solve these problems, nor to continue over a longer time frame. Local control has provided an opportunity to introduce new programs and concepts that may be a better reflection of local needs and the local cultural context. However, for the most part, at least in the realm of special education, the locally controlled schools do not seem to have taken advantage of this opportunity.

In some ways, local control has increased the difficulty of implementing new or specialized programs by isolating communities to a greater extent than they were under federal administration. It is unrealistic to expect small, isolated school systems to undertake educational program development without some type of support or assistance from outside the community. Moreover, locally controlled educational systems are more subject to local political pressures than those operated on a regional or province-wide basis. Given the special education evaluations summarized above, it is evident that not all the money earmarked for special education purposes was used for these purposes in some of the schools evaluated. Indeed, Indian education funds in general are not always protected from the demands of local band councils that they be used for other purposes (Hall, 1992).

What, then, should be done to provide more effective special education services in band-run schools? Based on the findings, we suggest the following five interrelated steps.

Develop a regional service network to support local programs. At present there does not appear to be an appropriate network of support services to assist individual schools, which are scattered in rural and northern areas throughout the province, to develop local policies, programs, and services. These support services must be able to provide specialized assessment, program development, and training assistance to local schools that require it. The provincial

system of support services in Manitoba provides an example of the kinds of needs that are to be met.

In addition, as shown in this article, Indian schools' special education programs are in need of monitoring and evaluation. Although a broad school evaluation once every five years is important, it is equally important that individual programs, including special education, receive regular feedback on their operation and performance. Because there is often a lack of local expertise in this area it would be useful to have a regional service that could provide such monitoring to those schools that require it.

A regional support network would also provide for sharing of information, program ideas, and services. Cooperation among communities is difficult under local control; a regional network would provide an avenue for cooperation.

Ensure that special education program funding fully addresses Indian needs. This review identifies some concerns over the funding of special education on reserves during the study period (1989-1992). First, although there seemed to be adequate provision for local special education services when the per capita grant and Level I funding were combined, schools seemed to have difficulty in fully accessing Level I funding. This may have contributed to the lack of development of various special education programs and services.

INAC has recently adopted a new per capita funding formula intended to cover both base funding and Level I funding. This new system should streamline the budgeting process and lend itself to improved local program planning. The new funding formula seems to be roughly equivalent to the provincial funding level, but it also seems to be a reduction in total funding available.

The local delivery system for special education needs to be developed. There are currently many inadequacies in the delivery of special education services in reserve schools. Although a regional support network would be helpful, the responsibility for developing these services rests with local Indian education authorities. In this respect there are three key factors:

1. Principals and directors of education must provide leadership in communicating the need for and purpose of special education in the school.
2. An educational program for staff and parents must take place, with the assistance of outside expertise where needed.
3. Special education funding must be protected from demands from other areas of educational programming or from politically based demands to use education funding for other purposes.

Provide resource teachers, and teaching staff generally, with inservice training and skill upgrading related to special education. Many resource teachers are not fully trained and experienced, and many other teaching staff are not clear about the role of the resource program and the way in which IEPs are developed and implemented jointly by classroom and resource teachers. Inservice training in this area is clearly needed, and in some cases resource teachers may need to be encouraged to enroll in more extensive professional upgrading.

Establish clear program standards and guidelines for special education in band-operated schools. Program standards or guidelines are needed to ensure that students in need of special education services have access to such services on reserves. These standards also need to be monitored, and the results of the

monitoring process should at least be reported to the community. There are three ways in which the establishment and monitoring of special education program standards could take place.

1. A province-wide Indian educational body could be formed by interested Indian education authorities to establish these standards and undertake the responsibility of monitoring them;
2. INAC could make special education funding conditional on spending the funds in specific ways, and back this up by requiring specific monitoring and reporting activities; or
3. INAC and/or a province-wide Indian educational body could contract with Manitoba Education and Training to monitor compliance with guidelines, which would be identical to provincial guidelines.

We note that the Educational Framework Agreement between the Assembly of Manitoba Chiefs and the government of Canada seeks to establish a province-wide educational body by 1995. This may provide the opportunity to create an effective system of supports and standards in the area of special education for reserve school systems in Manitoba.

Notes

1. The term *Indian* is used here to refer to the status or registered Indian population, whereas the term *Aboriginal* is meant to be inclusive of status and non-status Indians, Metis, and Inuit.
2. Manitoba withdrawal rates among Indian students may be higher than those of Indian students in other parts of Canada. Hull (1987, p. 44) found that there were higher rates of school leavers among Indian students in the western provinces than in central or eastern Canada, with Manitoba's student leaver rate being third highest among Canadian provinces.
3. These evaluations were conducted by Eleoussa Polyzoi, one of the co-authors of this article. Dr. Polyzoi's specific role was to examine and evaluate the special education programs at these schools. On average she was on site for approximately four to five days for each evaluation as a member of a team.
4. As of the 1994-1994 fiscal year INAC has instituted a new funding formula that provides a $400 per capita grant taking the place of the per capita grant base funding plus Level I funding. In other words, individual schools are not required to undertake an assessment process to justify Level I funding. The $400 is composed of $281 for base funding (a 2.5% increase over the 1993-1994 funding level) plus $119 for Level I funding. If 10% of the student population needed Level I support, the cost under the old funding system would have been $456 per student. If only 5% of students needed Level I support, the cost would have been $228 per student. The new funding formula, therefore, appears to be a substantial reduction in overall funding for special education.

References

Alberta Education. (1977). *The Alberta special education study: Condensed report*. Edmonton, AB: Alberta Education, Planning and Research.

Assembly of First Nations. (1988). *Tradition and education: Towards a vision of our future* (vol. 1). Ottawa, ON: Author.

Assembly of Manitoba Chiefs. (1984). *Our tomorrows today: Different paths to quality Indian education in Manitoba*. Winnipeg, MB: Author.

Assembly of Manitoba Chiefs. (1991). *Indian education funding study* (Final Report). Winnipeg, MB: Author.

Canadian Council for Exceptional Children. (1989, October). *Keeping in Touch: A Quarterly Newsletter from the Canadian CEC Office*. North Bay, ON: Nippissing University College.

Canadian Council for Exceptional Children. (1992, December). Provincial/territorial survey of services to exceptional children. *Keeping in Touch: A Quarterly Newsletter from the Canadian CEC Office*.

Council of Ministers of Education. (1983). *Survey of special education in Canada 1982-1983*. Toronto, ON: Author.

Education and Research. (1985). *A study team report to the task force on program review*. Ottawa: Supply and Services Canada.

Hall, D.R. (1992). FED-BOS: The federally controlled band operated school and the no-policy policy. *Canadian Journal of Native Education, 19*, 67-66.

Hull, J. (1987a). *An overview of the educational characteristics of registered Indians in Canada*. Ottawa, ON: Indian and Northern Affairs Canada.

Hull, J. (1987b). *An overview of Registered Indian conditions in Manitoba*. Ottawa, ON: Indian and Northern Affairs Canada.

Hull, J. (1990). Socioeconomic status and Native education in Canada. *Canadian Journal of Native Education, 17*(1), 1-14.

Indian and Northern Affairs Canada (INAC). (1987). *Classification and housing economic categorization of Indian bands by zone*. Ottawa, ON: Author.

Indian and Northern Affairs Canada (INAC). (1992a). *Basic departmental data—1992*. Ottawa, ON: Author.

Indian and Northern Affairs Canada (INAC). (1992b). *Indian register population by sex and residence, 1991*. Ottawa, ON: Author.

Island Lake Tribal Council. (1987, March). *Report on special education in the Island Lake Region.* Winnipeg, MB: Author.

Koens, P.R. (1989). Local control within a model of the sociology of knowledge: A study in curriculum. *Canadian Journal of Native Education, 16*(1), 37-44.

Manitoba Education and Training. (1989). *Special education in Manitoba: Policy and procedural guidelines for the education of students with special needs in the public school system*. Winnipeg, MB: Author.

Manitoba Education and Training. (1992). *Funding of schools: 1992/93 school year*. Winnipeg, MB: Author.

Mariash, L. (1989). *School evaluations: The effective schools cycle*. Winnipeg, MB: Indian and Northern Affairs Canada, School Evaluations/Education Planning.

McBride, C., Gagne, M., & Atwell, K. (1990). *Characteristics of public administration employment on-reserve; 1986 Census*. Ottawa, ON: Indian and Northern Affairs Canada.

McKnight, B. (July 14, 1988). *Letter from Bill McKnight, Minister of Indian Affairs and Northern Development, to all Indian bands and councils with attached Policy for the Provision of Education Facilities and Outline of the School Space Accommodation Standards.* Unpublished document.

Morris, A. (1862-1979). *The treaties of Canada with the Indians of Manitoba and the North-West Territories*. Toronto, ON: Coles.

National Indian Brotherhood. (1972). *Indian control of Indian education* (Policy paper presented to the Minister of Indian Affairs and Northern Development). Ottawa, ON: Author.

Paquette, J. (1986). *Aboriginal self-government and education in Canada* (Aboriginal Peoples and Constitutional Reform, Background Paper Number 10). Kingston, ON: Queens University, Institute of Intergovernmental Relations.

Phillips, R.S., & Cranwell, C.R. (1988). The need for community-based special education programs in band-operated schools of Manitoba. In A.A. Ortiz & B.A. Ramirez (Eds.), *Schools and the culturally diverse exceptional student: Promising practices and future directions* (pp. 112-119). Reston, VA: Council for Exceptional Children.

Roberts, C.A., & M.D. Lazure. (1970). *One million children: A national study of Canadian children with emotional and learning disorders*. Toronto, ON: Crainford.

Stevens, H. (Principal Investigator). (1982). *A review of changes in the living conditions of the registered Indian population of Manitoba during the 1970's*. Winnipeg, MB: Social Planning Council of Winnipeg.

Working Margins Consulting Group. (1989). *Evaluation of Indian counselling services in Manitoba: Final report*. Unpublished report prepared for a steering committee of Manitoba tribal councils and Indian and Northern Affairs, Manitoba Region.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, March 1995, 63-83*

Grace V. Malicky

and

Charles A. Norman
University of Alberta

Perceptions of Literacy and Adult Literacy Programs

There are numerous definitions of literacy in the literature but relatively few data on the views of adult learners in literacy programs regarding literacy and even fewer on the views of adult literacy instructors. This study examined the views of 94 learners and 31 teachers regarding literacy as well as the actual classroom experiences of a subgroup of learners. Results showed that learners tended to view literacy from a fundamental perspective but entered programs for job-related reasons. Their teachers viewed literacy from a functional perspective but presented programs that were fundamental in nature. Emancipatory views were reflected to a very limited extent in either views of literacy or actual classroom experiences.

Bien qu'il y ait de nombreuses définitions de l'alphabétisation citées dans plusieurs documentations, il existe peu de données des opinions qu'ont les apprenant(e)s adultes participant aux programmes d'alphabétisation en ce qui concerne l'alphabétisation comme telle. Il existe encore moins de données sur les opinions qu'ont ces gens de leurs instructeurs et instructrices. Cette étude examine les opinions de 94 apprenant(e)s et 31 instructeurs et instructrices en ce qui concerne l'alphabétisation ainsi que les expériences en classe d'un sous-groupe d'apprenant(e)s. Les résultats indiquent que les apprenant(e)s considèrent l'alphabétisation d'une perspective fondamentale mais s'inscrivaient dans ces programmes pour des raisons reliées au marché du travail. De leurs côtés, les instructeurs et instructrices percevaient l'alphabétisation d'une perspective plutôt fonctionnelle mais présentaient des programmes de façon fondamentale en sorte. Des opinions émancipées ont été exprimées d'une façon très limitée lorsqu'il s'agissait soit des expériences en salle de classe, soit des points de vue concernant l'alphabétisation.

Introduction

Although some information is available on the concepts of reading held by adults in literacy programs (Gambrell & Heathington, 1981; Norman & Malicky, 1986), little research has been done on how instructors in adult literacy programs view literacy or on the congruence of views of literacy learners and teachers. In light of dropout rates from adult literacy programs that have been estimated to be as high as 70% (Quigley, 1992), it is important to understand not only the degree of consistency between learners and instructors in how they view literacy, but also the degree of consistency of their views with

Grace Malicky is a professor in the Department of Elementary Education and Director of the Reading and Language Center. Areas of research interest include emergent literacy, clinical reading, and adult literacy.
Charles Norman is a professor in the Department of Educational Psychology. Areas of research interest include emergent literacy, cognitive development, and adult literacy.

literacy practices in classrooms. Street (1992) stresses the importance of understanding the nature of the relationship between ideas and practices. He believes that literacy is a social, political phenomenon that always involves power relations.

There is considerable support in the literature for a sociopolitical view of literacy (Fingeret, 1992; Harman, 1987; Scribner, 1984). Kazemek (1988), for example, describes literacy as a personal and social process of coming to know that is political, cultural, and context-dependent. He decries current adult literacy programs that employ teacher-controlled, didactic methods and argues instead that "literacy is an ethical endeavor that has as its goal the liberation of people for intelligent, meaningful, and humane action upon the world" (p. 467). He would support a position where one first determines the goals of literacy education and then designs a program consistent with these goals.

There is often considerable difference in the backgrounds of adults in literacy programs and of their teachers. In general, learners tend to come from more disadvantaged backgrounds than their teachers (Hunter & Harman, 1979). In a study of the literacy practices in three communities in Piedmont, Carolina, Heath (1983) found that the programs and practices used in schools were often at variance with the cultural backgrounds and values in people's lives. When teachers modified programs to make them more consistent with the language and literacy backgrounds of the children in their classes, achievement levels of children from disadvantaged backgrounds increased.

A major purpose of the study reported in this article was to examine the perceptions of both learners and teachers regarding the nature and goals of literacy as well as their views regarding the literacy programs in which they were involved as teachers and learners. The focus of this part of the study was on the degree of congruence between the views of learners and teachers. Lack of agreement in perceptions of literacy and literacy programs could be one factor in high dropout rates from adult literacy programs. A second purpose of the study was to observe the nature of the experiences of a subgroup of learners in literacy classes to determine the extent to which programs reflected the goals and views of learners and teachers. A discrepancy between stated goals and the means to realize them could negatively affect the success of literacy programs. Analysis of data was conducted from the perspective of the following three major concepts of literacy that are widely reflected in the literature and have dominated the work of UNESCO from 1945 to the present (Lind & Johnston, 1986).

Concepts of Literacy

In 1946 the concept of fundamental education was adopted by UNESCO with the core content of adult education programs involving skills of thinking, speaking, listening, calculating, reading, and writing. Illiteracy was viewed as a disease to be eradicated and the major tool was primary school education. Adult literacy education was viewed as an extension of schooling and often the same materials and techniques were used. Few writers define literacy this narrowly today, although fundamental literacy is sometimes included as one aspect. For example, Valentine (1986) uses the term *general* literacy to refer to the reading and writing aspects of literacy and contrasts this with *functional* literacy. As Kazemek (1988) points out, this view of literacy is still reflected in

the instructional practices of many literacy programs. For the purposes of this study, fundamental literacy refers to the skills of reading and writing.

The concept of functional literacy began to influence international thinking in 1956 with Gray's survey of reading and writing conducted for UNESCO. Whereas Gray stressed the relative nature of literacy and the need to relate literacy training to the context in which individuals live, UNESCO linked the concept of functional literacy with social and economic development. The major focus was on literacy for work and increased productivity. Literacy was not viewed as an end in itself, but rather as a way of preparing people for social, civic, and economic roles. Literacy was viewed from a human capital paradigm of economics in which education is viewed both as a form of consumption and a form of investment. The goal of education from this perspective was to enhance human capital, which would in turn increase productivity and yield a positive rate of return. Although this narrow view of functional literacy is still reflected in some programs, many writers view functional literacy from a much broader perspective. For example, Valentine (1986) defines it as "an individual's reading and writing ability in relation to the reading and writing tasks imposed by, or existing in, the environment in which that individual resides and seeks to function" (p. 109). Levine (1982) focuses more on information when he defines functional literacy as "the possession of, or access to, the competencies and information required to accomplish those transactions entailing reading and writing in which an individual wishes--or is compelled—to engage" (pp. 263-264). In this study functional literacy is viewed from the broader perspective of Valentine and Levine, although it is important to note that one of contexts in which people deal with written language involves work.

Although functional literacy continues to dominate much of the literature in North America, UNESCO began to move away from a functional definition in 1975. The Declaration of Persepolis adopted at the International Symposium for Literacy in 1975 retained the relative and contextual aspects of literacy but stressed the political, human and cultural aspects as well. Literacy was defined as

> not just the process of learning the skills of reading, writing and arithmetic ... Literacy creates the conditions for the acquisition of a critical consciousness of the contradictions of society in which man [sic] lives and of its aims; it also stimulates initiative and his participation in the creation of projects capable of acting upon the world, of transforming it. (Hamadache & Martin, 1986, pp. 128-129)

This is the definition used in this study for the terms liberation or emancipatory literacy. From this perspective, literacy is not viewed as neutral, "for the act of revealing social reality in order to transform it, or of concealing it in order to preserve it, is political (Hamadache & Martin, 1986, p. 128). This liberation or emancipatory view of literacy reflects the work of Freire (1970) and others (Hunter, 1982; Kretovics, 1985; Stanage, 1986) who have focused on empowerment and the ability to bring about change in inequities and injustices in society.

Generally, the goal of literacy programs based on functional and fundamental views of literacy has been to adapt individuals to fit into the existing

sociopolitical order and hence to preserve this order. The goal of liberation or emancipatory literacy, on the other hand, is for individuals to act on society and bring about change.

Procedures

The Sample and Context of the Study

The sample of learners consisted of 94 adults enrolled in literacy programs in one urban center. The gender, age, status, and mean reading level of subjects are presented in Table 1 along with the type of literacy programs in which subjects were enrolled. The range of reading grade scores on the Tests of Adult Basic Education (TABE, 1987) was from level 2.1 to 9.2. The higher proportion of females to males was reflective of the population in adult basic education programs in the area as was the relatively high proportion of immigrants.

A subgroup of 18 learners was randomly selected for more in-depth study, and the teachers of these learners were also involved in the research. Half the subjects in this subsample were immigrants and the other half Canadian born. Twelve were female and six were male. Eight of these subjects remained in adult basic education, bridge (a transition between ABE and high school classes for ESL students) or high school classes across the duration of the study, whereas the others either went to different types of programs or left for personal reasons (e.g., health, family, etc.). Of the five who went to other programs, four went into trades or pretrades programs whereas the other went to a high school program in another setting because of dissatisfaction with his previous setting.

Most instructors of adult basic education, bridge, and high school programs who taught the 18 subjects in the subsample were interviewed to obtain their views of literacy and to describe typical literacy classes. Views on literacy were

Table 1
The Sample of Literacy Learners

Variable	*Number or Mean*
Gender	
Females	61
Males	33
Mean Age	29.5
Status	
Canadian-born	40
Immigrants	54
Attendance in the Literacy Program	
Full-time	69
Part-time	25
Nature of the Literacy Program	
Formal	84
Non-formal	25
Mean Reading Level on TABE	5.0

obtained from 26 instructors, 19 of whom were teachers in adult basic education or bridge classes and seven of whom were in high school programs. Five additional instructors provided descriptions of typical literacy classes.

Data Collection and Analysis

Data were collected from the sample of 94 adults attending literacy programs and from 31 teachers of these programs. Both formal testing and interviews were used to collect information from learners. The Test of Adult Basic Education (TABE) was used as a screening tool to select learners below a high school reading level because the major focus of the study was on adult basic education. All learners were interviewed at the beginning of the study using a structured interview schedule to obtain demographic data as well as information about their concepts of literacy and reasons for entering the literacy program. This interview schedule was based on a questionnaire used by Davis and O'Brien (1985) with adults in literacy programs in Nova Scotia. The purpose of this interview was to describe the adult participants in literacy programs as well as to identify the meanings they associated with literacy. The major changes to the interview schedule involved omitting some questions because Davis and O'Brien had noted that it took too long to administer. The initial interviews were audiotaped and responses were recorded on the interview sheets.

Across the three-year time period of the study learners were also interviewed at six-month intervals while they attended literacy programs and at the time they exited literacy programs (if they did). One focus of these interviews was on what they viewed as helpful in the literacy program and what else they would like to have seen in the programs. Two open-ended questions were used to elicit these comments: What do you think are the strengths of what you are doing in your classes? and What else would you like to do that you are not now doing in your classes? Although change across time was not of major concern in relation to these questions, the three-year time period of the study was essential to understand the relationship of participation in literacy programs with employment (Malicky & Norman, 1994b) and to provide data on more than one class for most learners in the study. Research assistants conducted interviews under the supervision of one of the principal researchers and all interviews were audiotaped and transcribed for analysis.

Learners in the subgroup of 18 were observed approximately every three months in their classes to obtain data on the nature of their actual classroom experiences. These observations were conducted for as long as the participants remained in literacy or high school upgrading classes across the three-year time frame of the study. Fieldnotes were kept of these observations, outlining what materials were used as well as providing a running commentary on what students and teachers did and said during the lessons.

The teachers of the subgroup of 18 were interviewed to determine their concepts of literacy as well as their perceptions of the literacy program they were providing. Questions that focused on ways in which literacy and literacy programs were perceived were selected from a questionnaire for literacy instructors developed by Davis and O'Brien (1985). Examples are: Why do you think being literate is important? What do you think constitutes being a literate person? What do you think should be the purpose of literacy programs like the

one in which you are involved? Would you describe a typical class or session for me? All interviews were audiotaped and transcribed.

Data collected from the total sample of 94 learners on the TABE and initial interview were quantified and tabulated where possible. Means were calculated to provide a description of the adults in the study. Responses to other questions such as literacy concepts and perceived strengths of programs were categorized and the number of responses falling into each category determined. Data gathered on follow-up interviews with learners, on interviews with instructors, and through classroom observations were read to identify emerging categories regarding perceptions of literacy programs by program participants, views of literacy held by instructors, content of programs and the nature of interactions in classrooms.

Learners' Perceptions of Literacy and Literacy Programs

The majority of learners were full-time students at the start of the study, attending classes during the daytime five days per week. Part-time students generally attended classes once or twice each week. Most of the adults in the study were from formal programs, with 51 attending literacy classes in a vocational center and the other 33 attending classes in a continuing education program sponsored by a school system. The remaining 10 adults were in nonformal literacy programs, five in a volunteer one-to-one program and the other five in community-based programs.

Funding for the programs involved in this study was provided by two government departments—Manpower and Education—with the result that these programs generally reflected one of the following two orientations: (a) a manpower orientation concerned primarily with enabling individuals to participate in the labor force, and (b) a social demand orientation concerned with enabling individuals to participate more fully in all aspects of society. In this study the literacy program situated in the vocational center had an explicit manpower orientation whereas the other programs had more of a social demand orientation.

Definitions of Literacy and Goals for Entering Literacy Programs

Program participants were asked directly to provide their definition of literacy during the initial interview (Are you familiar with the word literacy? What do you think it means?). Many were unfamiliar with the term, particularly those with English as a second language. Of the 36 who did respond, 25 or 69% included reference to reading, writing and/or literature in their definitions. Five referred to knowledge, five to education, schooling or learning, and one to English as a second language. Generally, definitions provided by participants tended to reflect a fundamental notion of literacy.

Further information on participants' views of literacy was obtained by asking them, Why did you enter this program? Many participants gave several reasons, but the most common were job-related: 83% believed that increased literacy could help them improve their job opportunities. In addition to job-related goals, several people cited personal/psychological reasons such as feeling better about themselves and developing self-confidence. These reasons were given more frequently by Canadian-born respondents (43%) than by immigrants (17%). Social reasons, such as meeting people or becoming more

independent, were given by approximately one quarter of the men and immigrant women, but 73% of Canadian-born women provided reasons in this category. This may reflect differences in men's and women's ways of knowing with women perceiving themselves more in terms of connections and relationships (Kazemek, 1988). Both general and specific educational goals were also important to the participants in this study. Immigrants (43%) gave more specific goals, such as learning how to read and improving English skills, than did Canadian-born respondents (16%).

Perceived Strengths of Literacy Programs and Suggested Changes

Participants were asked at regular intervals to comment on their perceptions of the programs they were attending. Some participants left the program before they could be given a follow-up interview, and hence the data presented in Table 2 were collected from 83 of the sample of 94 learners. Because some participants moved from one program to another during the study, comments on both basic literacy and high school upgrading programs are categorized in Table 2. The scores in the table indicate the number of comments received regarding areas that participants felt were strengths or in need of change at some point during their participation in programs.

Areas of Language Arts

The information presented in Table 2 indicates that participants frequently referred to traditional areas of language arts education as both strengths of programs and areas in which more instruction was needed, although this was less true for Canadian-born males than for the other program participants. The areas that participants perceived as most helpful were reading and writing, with 39% citing writing as a strength of their programs and 28% citing reading. The other two areas mentioned frequently were grammar (24%) and vocabulary (20%). It is not surprising that these two areas were viewed as valuable by immigrants because most were learning English as a second language, but it is interesting that several Canadian-born women (23%) included these areas as well. One woman talked about improving her language competence because she had a young child, and it may be that some of the other women also felt the need to present better language models for their children. In any case, this finding is consistent with the relatively high number of social goals given by

Table 2
Perceptions of Literacy and Upgrading Programs: By Learners

Focus of Comment	*Number of Comments Received*	
	Strength	*Area of Suggested Change*
Language arts areas	121	57
Specific instructional materials	5	0
Content	6	9
Specific activities	7	6
Interpersonal aspects	33	15
Organization of classes	48	26
Level/Amount/Pacing	3	26

Canadian-born women for entering literacy programs. None of the Canadian-born men cited vocabulary or grammar as strengths of or needs in their programs.

When asked what changes they wanted in their programs, women tended to provide more responses than did men, but even the women gave relatively few responses to this question. In relation to areas of language arts, two areas mentioned by several of the immigrant females involved more focus on speaking (22%) and listening (16%), again consistent with expectations for ESL learners. These areas were mentioned by only two of the Canadian-born participants. More reading, writing, spelling, and pronunciation were also mentioned by a small number of both immigrants and Canadian-born participants.

The relatively heavy focus on traditional areas of language arts education is consistent with definitions of literacy provided by program participants and confirms that a fundamental or traditional view of literacy was widely held. Hence it appears that participants brought this view with them to their literacy programs, although persistence of the view could also reflect the nature of the programs in which they were enrolled.

Content, Materials, and Instructional Activities

In general, skill areas such as reading and writing were mentioned much more frequently than content or topics. Only six participants identified general or specific content as a strength of their programs (e.g., the environment, Meech Lake, different life styles) and six noted general or specific content that they would like to have had included (e.g., books on animals, news, history of different nations). Three other participants cited problems with their programs related to content: two talked about content being useless or irrelevant and one identified a specific workshop on how to dress as a waste of time. Overall, though, content was not a strong focus of the program participants in this study, and there was little difference between immigrants and Canadian-born participants or between males and females in the extent to which they identified content or topics as strengths or needed changes in their programs.

Similarly, specific instructional materials or activities were rarely mentioned as either strengths or weaknesses of programs. A few respondents referred to specific books being used in their classes as good and to activities such as story projects and research papers as useful. They offered no suggestions regarding what materials they would like to use and tended to identify activities they didn't like rather than make more constructive suggestions. This is perhaps not surprising in that they would not have been aware of the range of options available.

Interpersonal Dimensions

Instead, learners frequently focused on interpersonal dimensions when talking about strengths and areas of needed change. Both immigrants (41%) and Canadian-born participants (53%) indicated that their teacher was a real strength of their program. Many simply indicated that their teachers were good, but others talked about what made their teachers good, for example,

> The teacher, I thought, was an excellent teacher. She knew what she was talking about and she wouldn't go any further unless everyone in the class understood it. (female, Canadian-born)

> I like the teachers ... they are real helpful and friendly and willing to help every time I ask them. (female, immigrant)

> All the teachers are really good. They really pushed me further than I ever thought I would go.... When I was at school before, they treated me like a child and I didn't get any benefit from it. Now they treat me like an adult and you know, you can talk to teachers like, when I was at school you didn't know the teacher's first name. (male, immigrant)

The somewhat higher percentage of Canadian-born than immigrant participants who cited their teacher as a strength of their program might reflect the heavier focus of Canadian-born respondents on personal/psychological goals for entering literacy programs. Interpersonal relationships would appear to be more critical to meeting these goals than would specific content or instructional activities.

It is important to note that not all comments about teachers were positive. Seven students felt that they needed more time or assistance from their teachers, and three others felt they were being treated like children. According to one woman,

> She teaches as if we're in an elementary class type of thing.... Like she, you know, it's like "Okay kids we're going to do this, this and this today." Like, I find that writing spelling words out is a real pain. I mean that's what grade ones and twoers do type of thing. (female, Canadian-born)

Two of the immigrants felt they were being treated unfairly by their teachers, for example,

> Sometime I think it maybe—prejudiced. I think the teachers ... I copy from the encyclopedia word by word, and they give me the paragraph back full of red marks. (female, immigrant)

Relatively few comments related to interpersonal relationships with other students. Some students such as one Canadian-born male talked about how he enjoyed "going to school and talking with the students," but generally the participants in this study viewed their teachers as being key to their literacy learning.

Organization of Classes and Pacing of Instruction

Program participants also commented on the way programs were structured, particularly those students in the computer-based literacy program. Initially, many were enthusiastic about working on computers (11 commented on this as a strength of their programs), but once the programs were underway, five people expressed concern that they were not really learning to use computers, for example,

> I really thought the computer program was gonna be like where you learn the computer, like doing it so you go out and get a job. (female, immigrant)

A need for class instruction along with use of computers was identified by three individuals and six indicated that they would like more class time. Typical of these comments were the following:

> It's fun working on the computers but if you wanted to sit down there and learn, I find you learn more on an individual classroom basis. (male, Canadian-born)

> I like the classroom teaching you know. In the classroom, I mean, there are more things, like a teacher teaching you and additional things that one might add and you get more knowledge from that. (female, immigrant)

Not only those participants in the computer-based literacy program expressed a need for more class time. Five participants expressed a desire to access pull-out classes for extra assistance.

Several comments were related to level, pacing, and amount of material covered. Two participants were pleased that they were challenged and one that the material was easy. However, several were concerned that the program moved too slowly (6), that the material was too simple (8), or that there was too much review (7). Twelve of these comments were received from immigrants, several of whom had attained more advanced levels of schooling in their countries of origin, but some Canadian-born participants also felt programs did not move quickly enough for them, for example,

> We've worked for weeks on these prepositional phrases and subordinate clauses and all these new words. And she's sort of drilled that into us and now it's becoming quite boring. If she doesn't move to some new material soon, I don't know how much more the class will put up with it. (female, Canadian-born)

Far fewer participants felt that programs moved too quickly (2), presented too much (2) or didn't provide enough review (1).

The Subsample

The subsample of 18 subjects was involved in more in-depth study, including interviews with their teachers and observations in their classrooms. The major purpose of this part of the study was to obtain an indication of the views of literacy held by literacy instructors and the nature of literacy programs provided. It was also conducted to determine the degree of consistency between teachers' and learners' beliefs and between what teachers said about literacy and what they actually did in their classrooms.

Eleven of the students in the subsample were attending literacy classes in the vocational center at the beginning of the study. Most of these students were immigrants and most had taken ESL courses at the center prior to entering adult basic education classes. Of those who were in basic literacy classes at the beginning of the study, two began in a level 3-4 classroom, one in level 4-5, three in level 6-7, and two in level 7-8. Two other students were in bridge classes for the first observation. Nearly all these students were observed in more than one level, and five eventually moved on to high school classes. There were strict attendance requirements in this program and movement from one level to the next was based on test performance.

Six of the students attended literacy classes in the school system-sponsored adult basic education program. Students were involved in both computer-based and classroom instruction. Two of the students moved into the high school program sponsored by the same school system and one went on to a pretrades program at the vocational center. The final student began her program in a community-based literacy program and then entered the high school program at the vocational center.

Instructors' Views of Literacy and Literate People

As indicated above, 26 instructors of the 18 adult learners in the subsample were interviewed to find out how they viewed literacy. They were asked to tell why they thought literacy was important, what they thought constituted being a literate person, and what they thought the purpose should be of literacy programs like the one in which they were involved. Most instructors provided multidimensional descriptions of literacy and of literate people.

The most common descriptions given by instructors involved functional aspects of literacy; a total of 44 responses fell into this category (several informants gave more than one functional aspect of literacy). Typical responses in this category follow:

> To make people more employable.
>
> Handle forms and most things you run into in most situations.
>
> Can read and write well enough to function in society.
>
> Function in daily activities.
>
> Being able to get along or succeed in day-to-day living.

Fundamental descriptions of literacy were also prevalent in the data; 27 of the responses fell into this category. Most involved a reference to a specific area or aspect of language arts, but several also involved references to schooling or academic achievement.

> Basic understanding of spelling, punctuation and basic paragraphs.
>
> Basic abilities for reading and understanding the written word.
>
> Literature, reading and writing.
>
> Preparation for [other programs named].
>
> Reading is not just decoding but enriching what a person already knows.
>
> Writing as a means of communicating.

A similar number of responses focused on the sociopolitical (11) as the personal, psychological (11) aspects of literacy. Examples of sociopolitical responses included: "if you're not literate, you're third class," "the illiterate let somebody else take control of their lives," "to make people who are not easily governed." Psychological aspects of literacy included: "able to enjoy themselves," "gives a sense of self-esteem," and "self-confidence." Five respondents recognized that literacy is relative, depending on both the person and context. One, for example, said "it depends on what the person wants to do." Finally, eight responses reflected a more global/general view of literacy, such as "want a future," "aspire to a better life," "important in every facet of life," and "related to being human." When views of literacy were examined in relation to whether teachers taught in adult basic education, bridge, or high school contexts, there was little difference in how literacy was viewed.

What is most striking is that all instructors interviewed for the study included some aspect of functional literacy in their descriptions of literacy or literate persons. Most also included a fundamental notion of literacy in their descriptions whereas less than one quarter referred to emancipatory or psycho-

logical outcomes of literacy. One would anticipate on the basis of these results that most literacy programs would reflect a somewhat equal emphasis on functional and fundamental aspects of literacy, whereas there would be far less of a focus on sociopolitical aspects or personal development.

Two types of information were collected to determine the degree of correspondence between views of literacy espoused by instructors and the programs they provided for students. First, they were asked to describe typical class lessons, and second, approximately every three months a lesson was observed and fieldnotes taken.

Descriptions of Typical Lessons

Instructors were asked to describe a typical class and the following probes were used if necessary: What are the books and materials you use? How much time do you spend on each area? Descriptions of typical lessons were obtained from 31 instructors (nine instructors provided more than one lesson description because they taught two or more different courses or levels). Nineteen instructors taught adult basic education classes (two of these also taught high school classes), four taught bridge classes (one of these also taught high school classes), and eight taught high school classes. Adult basic education and bridge classes in the vocational center were frequently taught by teams of two instructors.

Most instructors included more than one area in their descriptions of typical classes, and these were categorized as shown in Table 3. In a level 7-8 class at the vocational center, for example, the teacher indicated that the students worked on reading in the morning and writing in the afternoon. Reading instruction was based on a workbook containing factual articles; students completed questions for homework, and these were corrected and discussed in class. Literature was also included in the program, with students reading stories and answering questions in class. Writing instruction was focused around two workbooks, one with an emphasis on grammar and the other on vocabulary.

In the adult basic education program sponsored by the school board, students spent a considerable portion of their time in computer-based literacy programs. In the Pathfinders program, students completed pretests and on the basis of their results the computer provided them with a list of materials to use. After completing the recommended exercises the students took a posttest on the computer. Teachers were available to answer questions both on the use of the computers and on what students were doing. Classes varied considerably across instructors; reading and writing were common components, with spelling, grammar, oral presentations, listening, and study skills included in various classes.

In the community-based literacy program the teacher included one half-hour every day on reading comprehension and also what she referred to as exercises on an area of need identified from the students' writing, for example, quotation marks. In addition, students worked for one and a half hours Mondays on writing, Wednesdays on reading comprehension, and Fridays on spelling. The teacher selected the materials for the students to read, and current topics served as the basis for brainstorming and writing.

Table 3
Aspects Included in Descriptions of Typical Classes

Aspects of language/literacy	*Percentage (number) of Instructors*					
	ABE (N=24)		*Bridge (N=7)*		*High School (N=21)*	
Grammar	33%	(8)	29%	(2)	29%	(6)
Vocabulary	19%	(4)				
Oral presentations	13%	(3)	14%	(1)	5%	(1)
Discussions	19%	(4)				
Reading						
Strategies	21%	(5)			10%	(2)
Literature	19%	(4)	14%	(1)	57%	(12)
Individual	21%	(5)	29%	(2)		
Question/Answer on reading	33%	(8)	43%	(3)		
Writing						
Skills	13%	(3)	43%	(3)		
Paragraphs/essays	58%	(14)	29%	(2)	33%	(7)
Journals	13%	(3)				
Study Skills	8%	(2)	14%	(1)		
Other	13%	(3)				

From the data in Table 3 and these few examples, it is apparent that in spite of the heavy emphasis on functional literacy evident in the definitions provided by instructors, descriptions of typical adult basic education classes focused almost exclusively on fundamental aspects. This was true to a large extent for high school classes as well, although there were some differences. The majority of lessons at the high school level included work on literature and writing, whereas those in adult basic education classes involved a wider range of aspects of language arts. Typical of programs provided at the high school level is the following description provided by a teacher in the vocational center setting. The program she offered included five major components: short stories, novels, plays, research projects, and eight book reports. The program was planned to meet the requirements of the provincial curriculum.

Although it is possible that instructors were aware of the discrepancy between their definitions of literacy and the programs they provided, only a small number of them commented on this inconsistency during interviews. One, for example, indicated that the goal of the program should be to increase employability and self-esteem but "instead we are preparing students for an academic program." Another expressed concern that some students

> leave here feeling like a failure because they haven't made progress academically. They still aren't any better off in terms of a job.... We may be doing a disservice by making people feel worse than they did when they came in.

She noted that "we are changing" but that students were not always in favor of this change.

> Sometimes when we give students the opportunity to talk about what will help them in their everyday lives, that's not what they want from us. They don't see education as providing that.

Both of these instructors wanted a heavier focus on functional aspects of literacy. A third instructor felt that the institution was focusing on both functional and fundamental aspects and she wanted to see a heavier emphasis on sociopolitical goals.

Classroom Observations

Data obtained through actual classroom observations were also analyzed to gain further insight into the degree of congruence between descriptions of literacy provided by instructors and what they actually did in their classrooms. These data were analyzed both in relation to the content of the lessons as well as the nature of interactions between teachers and students. The number of lessons observed in adult basic education settings was 42, the number in bridge settings seven and the number of high school classes 21.

Content

Table 4 provides an indication of what the content focus was in the lessons observed. Again, most of what was observed in classrooms could easily be classified according to traditional areas of language arts education. Students were frequently involved in reading and writing activities, although the nature of these activities varied depending on whether they were attending adult basic education, bridge, or high school classes. Students generally read a wider range of materials in adult basic education classrooms than in high school English classes where the focus was primarily on literature. Writing activities were similar across the two contexts, with the exception of journal writing, which was not observed in high school classrooms. Very few bridge classes were observed, so it is difficult to draw any generalizations about what was happening in those classes. It does appear, however, that there was a heavy focus on grammar in these classes probably because of the heavy enrollment of ESL students. In some of the bridge classes observed students spent the entire time working through a grammar book, completing exercises and correcting them together in class.

Teachers generally selected what students would read and assigned topics for writing (although for writing teachers sometimes gave several possible topics for students to choose from and left the possibility open for students to choose one of their own instead). Journal writing and independent reading opportunities did provide for some degree of student control over content, but these were far more common activities in adult basic education than in high school classrooms.

What Table 4 masks is that functional or emancipatory goals may have been served in some of the reading and writing activities. Short stories occasionally lead to heated discussions about topics such as child abuse or drunk drivers. Suggested writing topics were also sometimes controversial in nature and designed to heighten consciousness about issues, for example, "Women should/should not have equal job opportunities." Generally, however, teachers reported little input from students into curriculum decisions and most

Table 4
Aspects of Lessons Observed in Various Contexts

Aspects of language/literacy	*Percentage (number) of Lessons*		
	ABE (N=42)	*Bridge (N=7)*	*High School (N=21)*
Grammar	29% (12)	100% (7)	19% (4)
Vocabulary	12% (5)	29% (2)	
Listening	7% (3)		
Oral presentations			5% (1)
Pronunciation	2% (1)	29% (2)	
Reading			
SRA Kits/Independent Reading	5% (2)	14% (1)	5% (1)
Literature	19% (8)	14% (1)	50% (10)
Newspaper	19% (8)		
Specific Skills	17% (7)		14% (3)
Writing			
Journals	12% (5)		
Paragraph/essay	12% (5)	14% (1)	29% (6)
Specific Genre	5% (2)		5% (1)
Specific skills	12% (5)		33% (7)
Life/Study skills	5% (2)		
Subject areas	10% (4)	29% (2)	

discussions were related to skill development rather than to issues in the life worlds of the students in the classes.

In general, there was considerable correspondence between what teachers said they did in typical classes and what actually went on in classrooms; activities in classrooms were primarily focused on fundamental aspects of literacy. Bridge teachers demonstrated a heavier emphasis on grammar than they reported, but there were few other striking differences between what teachers reported and what they did.

Interactions

Classroom observations were also analyzed in relation to the nature of the interactions that occurred between students and teachers, and the results of this analysis are presented in Table 5. The types of interactions placed in the "other" category included such things as the teacher giving students words and asking them to find corresponding pictures, teachers writing questions on the blackboard and having students answer them orally, and the teacher reading words and having students repeat them. What Table 5 fails to capture is the sequence of interactions in classrooms. In order to provide some indication of this, observations of selected classes for two of the students are briefly described.

Sandra was observed first in a community-based literacy program and then moved to a high school program in the vocational center. At the beginning of the first class observed in the community-based literacy program (March), students were working at circular tables on reading comprehension exercises

Table 5
Interactions Apparent in Classroom Observations

Nature of Interaction	Percentage (number) of lessons ABE (N=42)		Bridge (N=7)		High School (N=21)	
Teacher lectures	19%	(8)	43%	(3)	33%	(7)
Teacher questions, students answer (oral)	33%	(14)			24%	(5)
Teacher reads, questions/students answer	29%	(12)			14%	(3)
Teacher/students do example together	19%	(8)	14%	(1)	29%	(6)
Teacher goes through test/assignment orally	14%	(6)			33%	(7)
Students do exercise independently, check orally in class	26%	(11)	43%	(3)	29%	(6)
Oral discussion	24%	(10)	29%	(2)	10%	(2)
Students read orally	5%	(2)	14%	(1)		
Students do exercise in groups	12%	(5)	29%	(2)	29%	(6)
Students work individually on assigned task	50%	(21)	14%	(1)	24%	(5)
Students do test	7%	(3)	29%	(2)	5%	(1)
Students work individually on task of own choice	2%	(1)			14%	(3)
Other	14%	(6)	14%	(1)		

while the teacher worked with one individual on a possessive pronoun exercise. The teacher then handed out an exercise on quotations. She put an example on the board and directed the students' attention to a page in the handout, asking them to tell what was wrong with the sentences. The students chimed in answers as the teacher read each sentence. On the next page she read the directions and asked the students to do three items individually. She then called on individual students to provide answers. The teacher moved to reading comprehension and explained that she wanted the students to generate a main idea statement for articles on garbage and then go through the 5W questions. She solicited the 5Ws from the class and wrote them on the board. She then gave a different newspaper article to each of three groups, checking with each group as they worked and commenting on the appropriateness of the statements they had written. Each group then shared the main idea they had written with the rest of the class, going through the 5Ws after they had done so.

The next fall Sandra was in a grade 10 English class in the vocational center. The tables were organized in rows facing the front of the room and a text entitled *College Writing Skills* was used during the lesson observed. The teacher asked students to turn to a particular exercise in the textbook and to do it. They were to rewrite a passage that had nine errors (run-ons, dangling modifiers, etc.) in it. The teacher circulated helping individual students as the others worked quietly on their own. When the students finished, they passed their papers to another student, and the teacher went through the sentences one by one, asking students if there were any mistakes. After checking to determine how many students got all the sentences right, the teacher assigned the next

exercise on subject-verb agreement in the text. The students completed it individually, passed the papers to another student and again marked it together. The teacher then said, "Let's give you a chance to talk for awhile." He asked the students to write two topics on two pieces of paper. Students pulled topics from a box and spoke in front of the class for 30 seconds. The teacher encouraged the speakers by saying, "Well done," "Good job," and the students rated each other on a 10-point scale.

The second student, Mary, began upgrading in the program sponsored by the school system and then enrolled in English 30 in the vocational center. During one class in the upgrading program (March) the teacher began by working through an exercise on prefixes and their meanings from a workbook entitled *Communication Skills*. The teacher then assigned an exercise on suffixes and the students completed it, helping each other figure out the directions and then working quietly. Mary frequently consulted the dictionary as she completed this assignment. The teacher went through the exercise with the students who chimed in with answers. He then gave them the following writing assignment on the board:

> Take one of the words you have worked with in lessons 5-8 and expand it to give it more meaning, that is understanding of it to your reader.
>
> OR
>
> Take one of your own.

The teacher explained the assignment and that the students would be editing their own writing. The students wrote until the end of class time.

From data in Table 4 and from these examples, it is evident that most classes revolved around activities selected by the teachers that students did and then frequently checked orally with the teacher. Most interactions among people were teacher-directed although students did do some teacher-assigned activities in groups and in four of the classes observed had an opportunity to work on a task of their own choice. Overall, however, teachers tended to define what counted as knowledge and had most of the power in classrooms. There were some exceptions as indicated in the following description of Mary in a high school English class at the vocational center one year after the class described above.

The class began by the teacher helping students fill in a form on short stories they had read. The form contained the headings title, author, genre, and plot. One student mentioned that Solchuk means "son of sun" and the teacher said, "See I told you guys. You teach me something new everyday." Another student asked whether Mr. Solchuk was the protagonist or antagonist. Rather than providing an answer, the teacher said, "That's a good question. What do you guys think?" After considerable discussion the class came up with a definition. Similarly, other questions were dealt with by the class rather than by the teacher. When discussing one of the stories, Mary commented, "You want to believe parents do things like this." She talked about conflict in her family with her mother and father and indicated that she hadn't mentioned this in class because she didn't know that the young people would understand. A male student responded, "I think we can empathize with you," and discussion ensued about lack of family support. Finally, the teacher brought the discussion back to the story by asking the class to think for tomorrow about the fact

that the father of the character in the story was coming around. As the students packed up to leave, another student came over to talk to Mary and said he thought they had had similar experiences. Mary later told the researcher that she never used to talk in class, but this teacher had done something different with this class—they worked in groups and talked about things.

What this teacher had done was give up some of the ownership of knowledge to the students in her classroom. Her definition of literacy was quite similar to that of many other teachers interviewed—she gave fundamental (reading, writing, listening, speaking, viewing), functional (dealing with everyday life, being able to use resources, where to get help), and personal, psychological (increased self-confidence) goals for literacy. However, in her description of a typical class, she referred to the transactive theory of Rosenblatt (1978) and indicated that there was a heavy focus on response to literature. Students were encouraged to talk and write about how they felt about the literature they read, bringing in their personal experiences. This was not the only class in which a teacher relinquished some of the control of content and classroom interactions to students, but it was rare rather than common among the upgrading classes observed during this study.

Conclusions and Discussion

Overall, the results of this study revealed considerable discrepancy within and between student and teacher groups in how literacy was viewed and practiced. When the learners entered adult literacy programs, they tended to hold a fundamental view of literacy, probably as a result of their past school experiences. However, they persisted in this view as they attended literacy classes in spite of the fact that when asked why they had entered literacy programs they frequently cited job-related (functional) goals. This would appear to reflect a mismatch between the perceptions of literacy held by adults in literacy programs and their goals for entering these programs. However, it is unlikely that the participants viewed this as a mismatch. Instead, they appeared to believe that a traditional education was the avenue to improved job opportunities in spite of the fact that there is little evidence in the literature to indicate that increased literacy will have a positive impact on employment (Graff, 1987; Levine, 1986). Indeed, most of the learners in this study returned to the same low paying, temporary jobs that they had held before entering literacy programs (Malicky & Norman, 1994b).

In contrast to program participants, literacy instructors tended to view literacy from both functional and fundamental perspectives, yet presented programs that were primarily fundamental in nature. Hence the programs they were providing were more consistent with the views of learners regarding the nature of literacy than with their own views. This seems to reflect a gap between theoretical orientation and practice in the classroom. Some instructors recognized the discrepancy between what they believed should be done and what they were doing, but generally felt constrained by their institutions and, according to one instructor, the students as well. Despite somewhat different mandates of the vocational, school system-sponsored and community-based programs, literacy content and instruction was markedly similar. The degree of consistency in the type of literacy instruction across the different contexts appeared to reflect the pervasiveness and impact of the fundamental view of

literacy instruction and learning in the province. To some extent this was determined in high school upgrading classes by the mandated provincial curriculum, but there is no mandated curriculum for adult basic education classes. In light of high dropout rates in the study (Malicky & Norman, 1994a), the strong emphasis on fundamental literacy needs to be critically examined in relation to goals and purposes of programs.

Emancipatory views of literacy were reflected to a very limited extent in the views of literacy instructors and students or in the content and interactions in classrooms. Classrooms tended to be teacher-dominated, with teachers determining the content of instruction and evaluating student performance. Students generally expressed positive comments regarding their teachers and although they valued opportunities to interact with other students they worked individually rather than cooperatively. Overall, there were few opportunities for students to share power with their teachers, and this did not appear to be an area of concern for most of them. The purpose of the literacy programs in this study could be viewed as helping students to adapt to or fit into the existing sociopolitical order, and hence reproducing this order (Apple, 1982; Giroux, 1983), rather than equipping them to act as change agents on society. It is important to note that there were classes in which students engaged in the kind of consciousness raising dialogue that Freire (1970) believes is an essential element of change. Hence it was possible to reach emancipatory goals even in high school upgrading classes with a mandated curriculum.

There were relatively few differences between females and males or between immigrants and Canadian-born program participants in their views of literacy or their comments regarding literacy programs. Immigrants tended to focus somewhat more heavily on the traditional areas and skills of language arts whereas Canadian-born subjects were more concerned with interpersonal dimensions of literacy instruction, but a relatively large proportion of comments from all groups fell into these two broad categories. There were the expected differences of immigrants valuing work on grammar, vocabulary, speaking, and listening, but even here Canadian-born females also identified vocabulary and grammar as significant areas of instruction. There were few differences based on gender or immigrant status on how content, materials, specific activities, classroom organization, and pacing of instruction were viewed.

A significant finding of the study involved the importance attributed by adult literacy students to their teachers as compared with materials or methods. Teachers spend a considerable amount of time searching for the "best" series or instructional technique, and publishers try to fill this need. The adults in this study viewed their teachers as a significant strength of their literacy programs and valued the assistance and explanations they received during classroom instruction. In contrast, they rarely referred to specific materials or techniques when describing strengths of their programs. Hence the results of this study do not support a heavy emphasis on methodology or technology. This was particularly evident in students' responses to the computer-based literacy program. Although some were enthusiastic about the program throughout the study, many expressed a desire for more interaction with teachers after the initial phase of their involvement in the program. It appears,

then, that more emphasis needs to be placed on teacher-learner interactions in planning and implementing literacy programs rather than on searching for the "best" methods or materials. In particular, the results do not support replacing teachers with computer-based instruction.

Although the finding that students viewed their teachers as crucial and generally "good" will be of comfort to literacy instructors, it raises a question regarding the independence of adult literacy learners. One of the basic principles of adult education is to lead adult learners in the direction of becoming independent and self-directed, and yet the learners in this study appeared to be quite dependent on their teachers, viewing them as necessary to learning. In a study of adult illiterates reentering the learning context, McDermott (1982) found that self-direction and independence increased slowly as adults experienced success and developed confidence in themselves as learners. Thistlethwaite (1983) also identified lack of confidence in ability to learn as an obstacle to self-direction and independence. Many of the students in the present study were lacking in confidence at entry to literacy programs (Malicky & Norman, 1994a). This does not mean that we should treat adults like children; rather it means that adults will require assistance to overcome their fears and anxieties and to develop confidence in their ability to learn. Adult literacy students require supportive, successful learning experiences to become independent, self-directed learners.

It is important to keep in mind when interpreting the results of this study that they are more a reflection of societal views of literacy and literacy teaching than of individual teachers. The teachers in this study appeared to be no more empowered than their students; they were constrained both by societal views of literacy as reflected in their students and by the institutions in which they worked. The institutions in turn were constrained by the views of policy makers and funders of adult literacy programs. There is a need for major partners in the adult literacy enterprise—learners, teachers, administrators, policy makers, funders—to critically examine their views of literacy and literacy learning. Only in this way will programs move beyond the current almost exclusive focus on fundamental literacy to achieve some of the emancipatory potential of literacy learning. Only in this way will literacy programs help people to make rather than take their place in society.

Acknowledgment

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

References

Apple, M.W. (1982). *Education and power*. Boston, MA: Ark Paperbacks.

Davis, N., & O'Brien, M. (1985). *Literacy programs and their participants in the Halifax-Metro area*. Halifax, NS: Mount Saint Vincent University.

Fingeret, H.A. (1992). Literacy in the USA: The present issues. In UNESCO Institute for Education (Ed.), *The future of literacy and the literacy of the future* (pp. 15-23). Hamburg, Germany: Editor.

Freire, P. (1970). *Pedagogy of the oppressed*. New York: Herder and Herder.

Gambrell, L.B., & Heathington, B.S. (1981). Adult disabled reader's metacognitive awareness about reading tasks and strategies. *Journal of Reading Behavior, 13*(3), 215-222.

Giroux, H.A. (1983). *Theory and resistance in education*. South Hadley, MA: Bergin & Garvey.

Graff, H.J. (1987). *The labyrinths of literacy: Reflections on literacy past and present*. London: Falmer Press.

Gray, W.S. (1956). *Maturity in reading: Its nature and appraisal.* Chicago, IL: University of Chicago Press.

Hamadache, A., & Martin, D. (1986). *Theory and practice of literacy work.* UNESCO.

Harman, D. (1987). *Illiteracy: A national dilemma.* New York: Cambridge Book Company.

Heath, S.B. (1983). *Ways with words: Language, life and work in communities and classrooms.* Cambridge: Cambridge University Press.

Hunter, C. (1982). Literacy for empowerment and social change. *Visible Language, 16*(2), 137-150.

Hunter, C.S., & Harman, D. (1979). *Adult illiteracy in the United States.* New York: McGraw-Hill.

Kazemek, F.E. (1988). Necessary changes: Professional involvement in adult literacy programs. *Harvard Educational Review, 58*, 464-487.

Kretovics, J.R. (1985). Critical literacy: Challenging the assumptions of mainstream educational theory. *Journal of Education, 167*(2), 50-62.

Levine, K. (1982). Functional literacy: Fond illusions and false economies. *Harvard Educational Review, 52*, 249-266.

Levine, K. (1986). *The social context of literacy.* London: Routledge & Kegan Paul.

Lind, A., & Johnston, A. (1986). *Adult literacy in the third world: A review of objectives and strategies.* Stockholm: Institute of International Education.

Malicky, G.V., & Norman, C.A. (1994a). Participation patterns in adult literacy programs. *Adult Basic Education, 4*(3), 144-156.

Malicky, G.V., & Norman, C.A. (1994b). Participation in adult literacy programs and employment. *Journal of Reading, 38*, 122-127.

McDermott, P. (1982). Conversational asides: The social context of an adult literacy class. *Journal of Reading Behavior, 14*, 461-474.

Norman, C., & Malicky, G. (1986). Literacy as a social phenomenon: Implications for instruction. *Lifelong Learning 9*(7), 12-15.

Quigley, A. (1992). Looking back in anger: The influences of schooling on illiterate adults. *Journal of Education, 174*(1), 104-121.

Rosenblatt, L.M. (1978). *The reader, the text, the poem: The transactional theory of the literary work.* Carbondale, IL: Southern Illinois University Press.

Scribner, S. (1984). Literacy in three metaphors. *American Journal of Education, 93*, 6-21.

Stanage, S.M. (1986). "Unrestraining" liberty: Adult education and the empowerment of persons. *Adult Education Quarterly, 36*(3), 123-129.

Street, B. (1992). Sociocultural dimensions of literacy: Literacy in an international context. In UNESCO Institute for Education (Ed.), *The future of literacy and the literacy of the future* (pp. 39-51). Hamburg, Germany: Author.

Tests of Adult Basic Education. (1987). Monterey, CA: CTB McGraw Hill.

Thistlethwaite, L. (1983). Teaching reading to the ABE student who cannot read. *Lifelong Learning, 7*, 5-7, 28.

Valentine, T. (1986). Adult functional literacy as a goal of instruction. *Adult Education Quarterly, 36*(2), 108-113.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, March 1995, 84-102*

Richard Conte

Jac J.W. Andrews

Melanie Loomer

and

Gillian Hutton
University of Calgary

A Classroom-based Social Skills Intervention for Children with Learning Disabilities

A social skills intervention consisting of coaching, role playing, and information sharing was implemented in a classroom for children who attended a special school for learning disabled children. Twelve children (10 males and two females, average age 12.2 years) participated in the experimental program that was implemented over a six-month period by a clinical psychologist who collaborated with the classroom teachers. Fifteen students (9 males and 6 females, average age 13.3 years) participated in the control group in which no formal social skills program was implemented. Relative to controls, children who participated in the experimental program showed superior sociometric status and improved social problem solving skills in interview situations involving responses to provocations. In addition, for one of these situations there was a significant relationship between change in social problem solving and change in sociometric data. The findings are discussed in terms of the determinants of the effects of social skills interventions.

Un programme d'intervention en habiletés sociales qui consistait à diriger la pratique de ces habiletés entre les élèves, à encourager les jeux de rôles, et à l'échange d'information fut implanté dans une salle de classe pour élèves ayant des difficultés d'apprentissage. Douze enfants (10 garçons et deux filles, âge moyen 12,2 ans) ont participé au programme expérimental qui a été implanté sur une période de six mois par un psychologue clinique en collaboration avec des enseignant(e)s en salle de classe. Quinze élèves (neuf garçons et six filles, âge moyen 13,3 ans) ont participé au groupe contrôle dans lequel aucune formation d'habiletés sociales formelle ne fut introduite. Relatif au contrôle, les enfants qui ont participé à ce programme expérimental ont démontré un statut supérieur sur l'échelle sociométrique ainsi qu'une amélioration dans leurs habiletés à résoudre des problèmes sociaux en situation d'interview. Dans ces interviews, on demandait des réponses fictives à des situations provo-

Richard Conte is in private practice and is adjunct associate professor of clinical psychology. His interests are in the area of social skills, attention deficit disorder, program evaluation, and parenting.
Jac Andrews is an associate professor in Educational Psychology. He is interested in cognitive strategy-based assessment and instruction, social skills intervention, issues in the academic and social development of children with mild disabilities, and the teaching-learning interaction in secondary and postsecondary education.
Melanie Loomer is a clinical psychologist at the Learning Centre. She is involved in assessment and counseling with children and adults with learning and attention problems.
Gillian Hutton is a research assistant at the Learning Centre and has interests in the areas of social functioning and cognitive strategy instruction.

catrices. En plus, pour une de ces situations il y avait un rapport significatif entre le changement dans l'habileté de résoudre des problèmes sociaux et le changement des données sociométriques. Les résultats sont discutés en termes de déterminants des effets des interventions des habiletés sociales.

Increasingly, researchers have become interested in links between learning disabilities and social skills difficulties (Baum, Duffelmeyer, & Geelan, 1988; Dudley-Marling & Edmiaston, 1985; Gresham, 1988; Gresham & Reschley, 1986; Wiener, 1987). This trend is reflected by the inclusion of social skills deficits in definitions of learning disabilities that have been proposed both by the Interagency Committee on Learning Disabilities (ICLD, 1987) and the Association for Children and Adults with Learning Disabilities (ACLD, 1985). Although these definitions have not obtained widespread acceptance and the status of social skills as a learning disability remains a controversial issue (Conte & Andrews, 1993; Forness & Kavale, 1991; Gresham & Elliott, 1989), there is little doubt that the social problems of children with learning disabilities need to be addressed (Andrews, 1989; Bryan, 1991; Schumaker & Hazel, 1984; Vaughn, 1985; Vaughn & LaGreca, 1988).

As detailed in the recent review of McIntosh, Vaughn, and Zaragoze (1991), social skills research with children who have learning disabilities has tended to focus on interventions outside of the classroom setting. In these pull-out studies children with social difficulties are targeted for intervention and then withdrawn from the classroom for the purpose of social skills training. Although such studies have been moderately successful in improving children's social behavior (Schneider, 1992), there has been a notable lack of success in transferring social skills from the training environment to the classroom setting (Gorney-Krupshaw, Atwater, Powell, & Morris, 1981; Schumaker & Ellis, 1982; Whang, Fawcett, & Matthews, 1984). The lack of transfer in previous pull-out studies has prompted us to implement a social skills intervention in classrooms and to provide training to the teachers so that they can cope more effectively with the social needs of their students.

Although there is an increasing trend to provide educational services to learning disabled children in regular classes, it is the case that many children, particularly those who are severely affected, receive their education in full-time learning disability placements. One of the potential disadvantages of full-time learning disability placement is that the students may be exposed to negative modeling of social behaviors that could conceivably lead to a deterioration in social behavior over time. In support of this view, Bierman, Miller, and Stabb (1987) found a decline in positive peer interactions over time in socially rejected boys in a special education setting. Bierman et al. (1987) suggested that over time rejected children may actually show a decline in the overall level of peer acceptance as a function of familiarity with each other. One purpose of the present study was to determine if a social skills program could improve overall peer status in an intact classroom of children with learning disabilities.

Even in special schools for students with learning disabilities, there are different ways to provide social skills training. On the one hand it is possible to identify individual students with social skills difficulties and to remove them from the classroom for the purpose of social skills training, which is the approach most commonly used in previous research (see McIntosh et al., 1991). In

contrast, in the present study we have opted to provide social skills training in the school classroom and to expose all children in the classroom to the same social skills program. There are several potential advantages to this in-class approach. Teaching social skills in the classroom increases the likelihood that awareness of appropriate social skills will become part of the fabric of the classroom. Appropriate social skills can be reinforced throughout the day, and not just during the one or two hours a week that children attend social skills classes. An in-class social skills program may also make it possible to provide the kind of extensive treatment that is needed in order to have impact on the social difficulties of many children with learning disabilities. In this regard McIntosh et al. (1991) highlighted the brevity of the intervention period as a factor that has limited the effectiveness of social skills interventions. By conducting an in-class social skills program, it should be possible to embed the intervention in the school curriculum.

Our approach in designing the present program was to utilize features of programs that had been successful in previous pull-out social skills interventions. Specifically, two aspects of previous programs appear to be particularly salient: the use of a combined coaching/role playing intervention and the use of information sharing techniques between peers. First, with regard to coaching and role playing, research on social isolates (Oden & Asher, 1977) and aggressive children (Bierman & Furman, 1984) suggests that a coaching/role playing model of intervention can lead to relatively long-term gains in social functioning. Oden and Asher (1977) describe three components in their coaching and role playing methodology: verbal explanation of social skills, opportunities to practice the skills by playing with a peer, and a postplay review session. In their work with social isolates, they found that this coaching procedure led to long-term improvement in peer sociometric status. Gresham and Nagle (1980), Ladd (1981), and Ladd and Mize (1983) have also shown that coaching/role playing intervention significantly increased the number of positive peer ratings received by rejected children.

In developing the information sharing method that was used as a component of the present program, Bierman (1986) and Bierman and Furman (1984) have stressed the importance of having students share information about themselves and about their feelings with their peers. Fox (1989) showed that having students with learning disabilities share information about each other resulted in greater improvements in peer acceptance than a program in which students exchanged information about academic subjects.

In addition to the coaching/role playing and information sharing, some components of the program were directed specifically at two areas of deficiency in children with learning disabilities: strategy utilization and metacognitive awareness. It has been shown that children with learning disabilities employ inefficient and ineffective strategies (Hallahan & Bryan, 1981; Hallahan & Reeve, 1980; Torgesen, 1980). For example, Havertape and Kass (1978) recorded the vocalized self-directions of students with learning disabilities and average achieving students as they were attempting to solve problems. Relative to average functioning children, children with learning disabilities applied fewer appropriate strategies to problem solutions. Indeed, much of the time they appeared to implement strategies that were random and impulsive. In the

social domain Carlson (1987) found deficiencies in the ability of children with learning disabilities to generate appropriate strategies in conflict situations relative to those displayed by children without learning disabilities.

With respect to the issue of metacognition, other studies indicate that many children with learning disabilities have difficulty choosing and deploying effective strategies because of deficits in metacognitive (executive) processing. For example, it has been shown that they cannot monitor their comprehension of new information and experience difficulties in regulating their learning by planning, checking, revising, and evaluating their problem solving activities (Bos & Filips, 1982; Pressley & Levin, 1987; Wong, Wong, Perry, & Sawatsky, 1986).

In the present research, strategy training was provided by teaching students a procedure for generating appropriate strategies to problematic social situations. Metacognitive awareness was encouraged by having students monitor and evaluate their solution strategies in a group problem solving format (more detail regarding these procedures is provided in the Method section below).

In the present research three major hypotheses were tested. First, if the experimental program was effective in improving social relationships, then it should lead to a climate of greater social acceptance in the classrooms in which the program was implemented. Thus participation in the social skills program that involved coaching, role playing, and information sharing would lead to improved sociometric status for the class as a whole where the program was implemented. Second, if the strategy training aspect of the program was effective, then students in the experimental group should generate more appropriate responses to novel hypothetical social problem solving situations than the students in the untreated group. This hypothesis was based on the assumption that repeated practice in generating strategies in response to problematic social situations would enable students to develop more effective problem solving strategies that would generalize to novel role-play situations. Finally, if strategy training was the aspect of the program that led to improvements in social acceptance in the classrooms where the program was implemented, then it seemed reasonable to suppose that change in appropriate strategy use would predict change in sociometric status. Accordingly, the third hypothesis was that treatment-induced change in strategy use in the novel role-play situations would predict treatment-induced change in social acceptance.

Method

Subjects

The subjects in this study were students who attended a special school for severely learning disabled students. The criteria for entrance into the school are that students show evidence of a significant discrepancy between intellectual potential and achievement and that they have had a previous unsuccessful experience in a special class placement for students with learning disabilities. School regulations prevented us from obtaining standardized IQ test scores; however, we did have access to the intellectual assessment reports. Analysis of these reports indicated that all the students in the study were in the average range of intellectual ability. In addition, all children in the study were at least two years below expected age level in reading as indicated on the annual individual reading assessments administered by each student's teacher. In

these assessments, informal reading inventories such as the Jerry Johns Basic Reading Inventory (Johns, 1985) and the Burns and Roe Informal Reading Inventory (Burns & Roe, 1985) were used.

Each class in the school has approximately eight students. Teachers of similar aged students tend to work collaboratively and combine their students throughout the day. For this reason, two classes of eight students participated in the experimental group and two classes of eight students served as controls. Because of the small size of the school it was not possible to find two groups of students of exactly the same age. For this reason there was a statistically significant difference ($t(26)=7.87$, $p<.01$) in the average age of students in the experimental group (10 males and 2 females; 12.2 years, $SD=.32$) as compared with students in the control group (9 males and 6 females; 13.3 years, $SD=.42$). Sample size differences between the two groups (experimental group $n=12$; control group $n=15$) were because parents of four children in the experimental group and one child in the control group did not return the consent form. The consent form described the general purposes of the study and the evaluation instruments to be used. Parents of children in the experimental group were informed that all students would participate in the social skills program and that consent for assessment only was being requested. Parents of children in the control group were informed of the purposes of the study and the fact that consent for assessment was being requested. Analysis of reading scores obtained from school-administered reading tests indicated the children in the experimental group were functioning at a 2.9 grade level whereas students in the control group were functioning at a 3.7 grade level. Although this difference was significant ($t(26)=2.64$, $p<.02$) a subsequent analysis using age as a covariate indicated that the difference in reading scores between groups was attributable to age differences; that is, when age was introduced as a covariate, the analysis was nonsignificant ($t(25)=1.21$, $p>.2$). Analysis of covariance (using age as the covariate) on the math scores indicated no significant difference between groups ($t(25)=.66$, $p>.5$).

Procedure

The treatment subjects received the social skills program twice a week for 45 minutes over the course of six months (January to June). The two classes that formed the experimental group were combined for the social skills lessons. The control group received no formal social skills program. Based on interviews with the control teachers, social difficulties were dealt with as they arose. The teachers indicated that they often dealt with social issues by presenting the issue to the class as a whole so that students could offer their views.

In this intervention study a collaborative consultation research model was used that allows individuals with diverse expertise to work together toward helping others (Idol, Paolucci-Whitcomb, & Nevin, 1986; West & Cannon, 1988; West & Idol, 1990). Consultation models are designed to prevent and remediate learning and behavior problems as well as coordinate intervention programs (West, Idol, & Cannon, 1989). According to this model, individuals who are working together should negotiate their responsibilities and expectations, define the nature and parameters of the problem, generate and select intervention recommendations, evaluate the intervention, and redesign the intervention if needed (West & Idol, 1990).

In this study a clinical psychologist worked with teachers in the classroom. The clinical psychologist and the teachers determined their roles and responsibilities with respect to the social skills intervention program, defined the nature of the students' social problems, and worked together to develop, modify, and evaluate the social skills program.

As noted previously, there were two main components of the program: coaching/role playing and information sharing. In the coaching/role playing aspect of the program, the students were presented with situations (see Table 1) that were jointly decided on by the clinical psychologist, teachers, and students. The procedure described by Oden and Asher (1977) was adapted to the whole class intervention. Each situation was described orally to the class and the students were encouraged to give their views on how they would react to the social problem. Once everyone's ideas had been aired, students broke up into small groups in order to develop role-plays that demonstrated the problem and how they would cope with it. The presentation of the role-plays was then followed by a debriefing discussion in which the students were encouraged to evaluate and monitor the strategies that had been demonstrated in the role-plays. In giving guidance to the students regarding their responses to the role-plays, three considerations influenced the teachers and the psychologist. First, the students should develop active responses to situations rather than passively going along with whatever other children were doing in the situation. Second, children were coached to stop and think before acting. Third, they were encouraged to generate strategies and arrive at compromise solutions that would take all viewpoints of participants into consideration.

The information sharing aspect of the program initially took the form of self-awareness activities (i.e., how I see myself; how others see me; what are my interests? what are my strengths and weaknesses?) that were eventually shared with other students. Subsequently, students were given the opportunity to share their perceptions of each other (e.g., students were given the assignment of saying something positive about other children in the class) and to compare their self-perceptions with the perceptions that were reported by other students. The self-awareness/information sharing aspect of the program was derived from the Metacognitive Approach to Social Skills Training (Sheinker & Sheinker, 1988). In implementing this aspect of the program it became evident that students experienced difficulties in exchanging information because of a style of communication that was often excessively confrontive. To address this issue the students were divided into three groups that were each led by an

Table 1
Situations Targeted for Intervention

1.	Knowing how to respond when the teacher asks a student why he/she is having difficulty reading.
2.	Knowing how to ask for directions when you get lost.
3.	How to approach a classmate who you would like to make friends with.
4.	How to ask a person about his/her beliefs or culture when they are different from your own.
5.	How to join a new club at a new school you are attending.
6.	How to cope with a new school setting.

adult (the clinical psychologist and two classroom teachers). They were provided with specific exercises and were directed toward discussing issues (such as the difference between a fact an opinion) that would facilitate communication and information sharing.

Evaluation Measures

Checklist data. It is well established that 30-40% of children with learning disabilities have hyperactivity (Lambert & Sandoval, 1980). This fact is particularly relevant to the present study because of evidence that measures of hyperactivity are good predictors of peer sociometric ratings (Bruck & Hebert, 1982). Thus, prior to the start of the study, teachers were asked to complete one copy of the Abbreviated Conners Questionnaire (ACQ; Conners, 1969) for each student. The ACQ is usually used as an indicator of hyperactive behavior, and the 10 items on the scale consist of characteristic behaviors of hyperactive children (e.g., fidgety behavior, short attention span, temper tantrums). Each item is rated on a scale of 0 to 3: not at all, just a little, pretty much, and very much. The scores on the Conners items were summed for each student and this score was used to provide an index of hyperactive behavior.

Teachers also completed a copy of the Taxonomy of Problem Situations (TOPS; Dodge, McClaskey, & Feldman, 1985) for each of their students. The TOPS consists of 44 items that describe common problematic social situations (e.g., name calling, difficulties in turn taking) in a school context. Each item is rated on a five-point scale with 1 signifying that the behavior is never a problem and 5 that it is almost always a problem. In filling out the TOPS, teachers were instructed to score an item as a problem in terms of the appropriateness or inappropriateness of the child's response to the problematic situation. For example, a child would be rated as having a problem with name calling only if the teacher believed his or her reaction to name calling was inappropriate. For each student the responses to the 44 TOPS items were summed in order to provide an overall index of each student's social difficulties.

Students were asked to complete a self-perception questionnaire in a group-administered format that was identical to the TOPS except that the questions were reworded (by changing the pronouns from the third person to the second person) to make them answerable by the students themselves. In this case, students were instructed to indicate how problematic it was for them to react to the situation. As with the original TOPS questionnaire, each item is rated on a five-point scale with 1 signifying that the behavior is never a problem and 5 that it is almost always a problem. The students' responses were averaged and a total score was calculated as an index of each student's perceived social difficulties. In addition, the items from the adapted TOPS for students were used to construct the hypothetical situations as described below.

Hypothetical situations. In order to measure their strategy use, students were asked to respond to a number of hypothetical problem solving situations during individually conducted interviews. The hypothetical situations were derived from the items that appeared on the student version of the TOPS. In developing the hypothetical situations, the six student TOPS items that posed the most difficulty across all students in the study were used. The six items included: name calling; having a friend refuse an offer to play, and then seeing

the same friend play with someone else a short time later; having a group of classmates laugh at you when you do poorly at a game or activity; having friends leave you out of an activity; asking a friend for an object that you want; having to ask for something back that a friend has borrowed from you. These items from the student version of the TOPS were reworded into hypothetical problem solving situations. For instance, the hypothetical situation that addressed the problem of name calling was worded as follows: "If you attempted to join a group of other students and one of them said 'What do you want egg-head,' how would you respond?" Another scenario asked how a student would react if he or she had lent a possession to another student and the student was late in giving it back. In this case the student was asked to generate the name of a valued object that might be lent to another student. He or she was then asked, "If you lent (insert name of object) to a friend and he/she was late in giving it back, what would you do?" Responses to these situations were coded by trained raters with respect to seven categories derived from Carlson (1987) described in Table 2.

The first three categories (assertive, accommodative, and evaluative) were considered positive responses in light of the program goals to be actively involved, to stop and think, and to arrive at compromise solutions. The last three (passive, egocentric, and antisocial) were classified as negative. Students were presented with the hypothetical situations in two different ways. First, they were presented with the situation and asked what they would do. After responding to this question, they were given the same situation with a specific goal in mind (e.g., in the name calling situation they were asked to indicate what they would do to get the other child to stop calling them a name).

Table 2
Hypothetical Problem Solving Scoring Categories

Positive Strategies

1. *Assertive*
 Responses that deal with the problem in a direct, forthright manner. They allow the child to get what he/she wants while fostering or maintaining good social relationships
2. *Accommodative*
 Responses that are less direct and self-assured. They may not result in the child getting what he/she wants but will still foster or maintain good social relationships.
3. *Evaluative*
 Responses in which the child reflects on how he/she would feel or think in the situation. These include responses in which the child tries to see the situation from the other child's perspective.

Negative Strategies

4. *Passive*
 Responses that do not deal with the problem or seek social contact with others. The child will ignore, leave, or "not worry about it."
5. *Egocentric*
 Responses in which the child seeks to get his/her way at the expense of social relationships. This includes lying, stealing, cheating, nagging, whining, threatening, taking over the game, being bossy or belligerent.
6. *Antisocial*
 Retaliatory responses. The child does not seek the goal and only wants to repay other(s) for some perceived injustice. The child wants to hurt others either emotionally or physically.

Raters (the second author and a bachelor's level research assistant) who scored the students' social problem solving responses required approximately 10 hours of training in order to achieve an interrater reliability of .85 or greater. Reliability was calculated by dividing agreements by agreements plus disagreements of ratings obtained from two independent raters on 10% of the interviews that were conducted. One of these reliabilities was calculated before formal scoring was initiated. The second was conducted after half the interviews had been scored. The reliability scores were 88.2% and 92.1% respectively.

Sociometric evaluation. Sociometric ratings were obtained by having students rate all other children in the two classes that received the experimental program. The two classes were separate for most of the day, but came together for approximately six hours per week (for social skills instruction and for some language arts periods). Thus approximately half of the students who were rated by each student were familiar because they were in the same class for most of the day. The other half of the students were less familiar because they were in the same class for only six hours per week. The ratings were conducted in the same manner in the control classes. In this case children from the two control classes came together for approximately six hours per week for health and language arts. It is also important to note that there were only about 80 students in the school in which the study took place, so most students, especially those of similar age, were quite familiar with one another. Sociometric ratings were obtained in individual interviews in which subjects were given a deck of cards with the name of one student on each card. They were asked to sort the cards in response to three questions: (a) how much would you like to sit next to this person on the school bus? (b) how much would you like to hang out with this person on the playground? and (c) how much would you like to work with this person on a class project? The student was asked to sort the name of each child into one of three piles that were labeled: a lot, a little, not at all. For each child the ratings given by all other children of him or her were assigned a numerical code (a lot=1; a little=2; and not at all=3), summed and then entered into a data base for statistical analysis. Thus popular children would be those with relatively low scores; unpopular children would be those with relatively high scores.

Results

Checklist Data

Analysis of the TOPS completed by teachers indicated eight situations where the average ratings of teachers exceeded 3.0 (out of a possible 5). Student self-ratings on the adapted TOPS for students were generally lower with only two situations exceeding 3.0. A preliminary analysis was undertaken to determine if there were pretest differences between the experimental and control groups, on the teacher and student TOPS. This analysis revealed that the average teacher TOPS ($t(26)=.64$. $p>.1$) and student TOPS ($t(26)=.25$, $p.>1$) did not differ between groups.

Analysis of the pretest Abbreviated Conners Questionnaire responses indicated no differences between groups ($t(26)=.28$, $p>.5$; experimental group $M=8.75$, $SD=7.12$; control group $M=8.13$, $SD=4.09$).

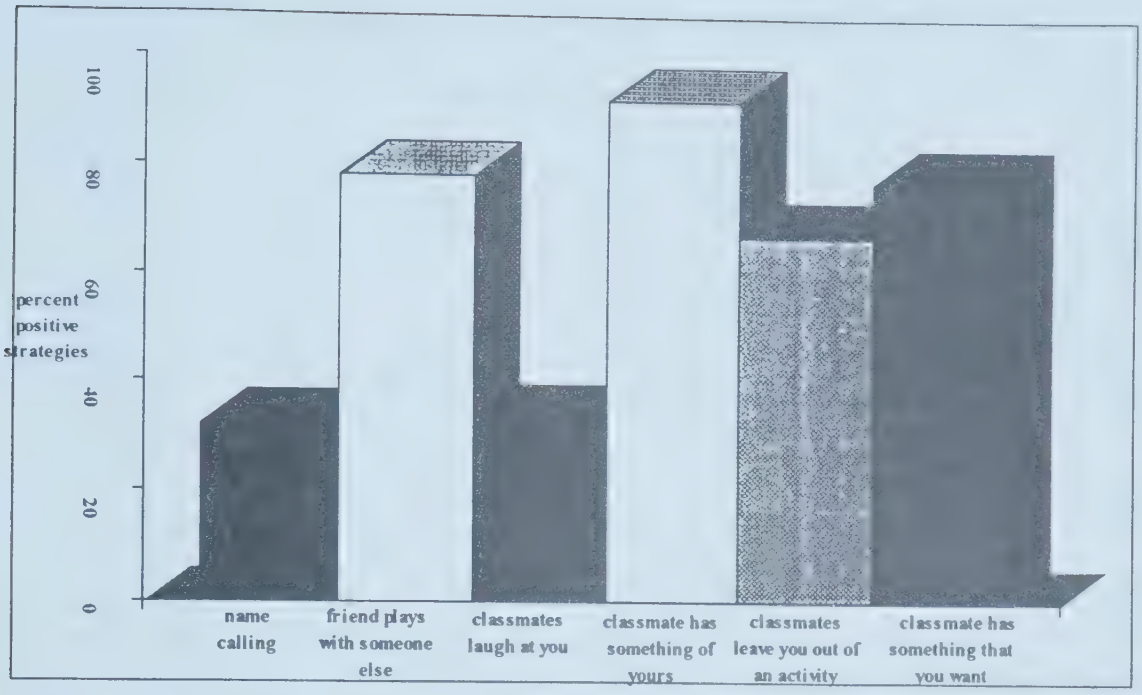


Figure 1. Percent positive strategies at pretest as a function of situational context.

Hypothetical Solution Strategies

As noted previously, students' categorized responses were reclassified into two solutions: assertive, accommodative, and evaluative categories were defined as positive solutions and passive, egocentric, and antisocial categories were defined as negative solutions. Pretest responses averaged across both experimental and control groups are shown in Figure 1, which indicates that there was considerable variation in the level of strategy use across situations. Two situations (name calling and being laughed at by classmates) showed a markedly lower incidence of positive strategies at pretest as compared with the other situations (χ^2 (5)=49.11, $p.<01$).

As noted previously, each hypothetical situation was administered in two ways. In the standard presentation condition, students were presented with the situation and then asked what they would do. In the goal-given condition, after the situation was presented students were given a social goal and then asked what they would do to achieve that goal. Between-group percent positive responses at both pre- and posttest to the hypothetical situations are shown in Table 3 for the standard presentation condition and in Table 4 for the goal-given condition. First, for the standard presentation condition, in order to determine the presence of pretreatment differences, between-group responses were compared using the Fisher Exact Probability Test. This analysis revealed that the responses of the experimental and control groups did not differ on any of the six situations. For the goal-given condition, there were pretreatment differences in favor of the experimental group for the situation in which "a peer has a toy, game, or object that I would like" ($p.<03$). Analysis of treatment effects were made in two different ways. First, posttest responses were compared using the Fisher Exact Probability Test. This analysis revealed that there were significant differences in only two situations: name calling, $p.<02$; and being laughed at by classmates in the goal given condition, $p.<03$. Second,

Table 3
Percent Positive Strategies in Responding to the Hypothetical Situations

	Experimental		Control	
	Pre	*Post*	*Pre*	*Post*
Name calling	25.0	60.7	40.0	21.4
Children laugh at you	36.4	54.5	33.3	40.0
You ask a child to play and find that they have played with someone else	73.3	60.0	83.3	66.7
Friends leave you out of an activity	58.3	58.3	72.7	90.9
A classmate has an object that you want	71.4	92.8	83.3	75.0
A friend has something of yours that you want back	93.3	86.7	90.9	90.9
Mean	59.6	68.8	67.2	64.1

analyses of the within-group change scores indicated significant positive change as assessed by a binomial test in the experimental group but not the control group for both of these situations: name calling ($p.<05$), and being laughed at by friends in the goal-given condition ($p<.05$). There were no significant change scores for any of the other situations in the experimental group and no significant effects for the control group for any of the hypothetical situations.

Sociometric Status

Change in sociometric status was assessed by assigning a score to each student based on the ratings received from the other students in the class (1=a lot; 2=a little; 3=not at all). Thus in this system a high score was indicative of a low status rating. The sociometric data are shown in Figure 2.

To evaluate sociometric scores a repeated measures split plot analysis of covariance with age as the covariate was performed with one between-group variable (GROUP: experimental vs. control) and two within-subject variables (SITUATION: with 3 levels: bus, playground, working on a class project and

Table 4
Percent Positive Strategies in Responding to the Hypothetical Situations in the Goal-given Condition

	Experimental		Control	
	Pre	*Post*	*Pre*	*Post*
Name calling	50.0	58.3	18.2	9.1
Children laugh at you	20.0	70.0	40.0	26.7
You ask a child to play and find that they have played with someone else	83.3	91.7	86.7	86.7
Friends leave you out of an activity	72.7	72.7	100.0	75.0
A classmate has an object that you want	100.0	90.0	58.3	83.3
A friend has something of yours that you want back	90.9	54.5	64.3	64.3
Mean	69.48	72.86	61.25	57.50

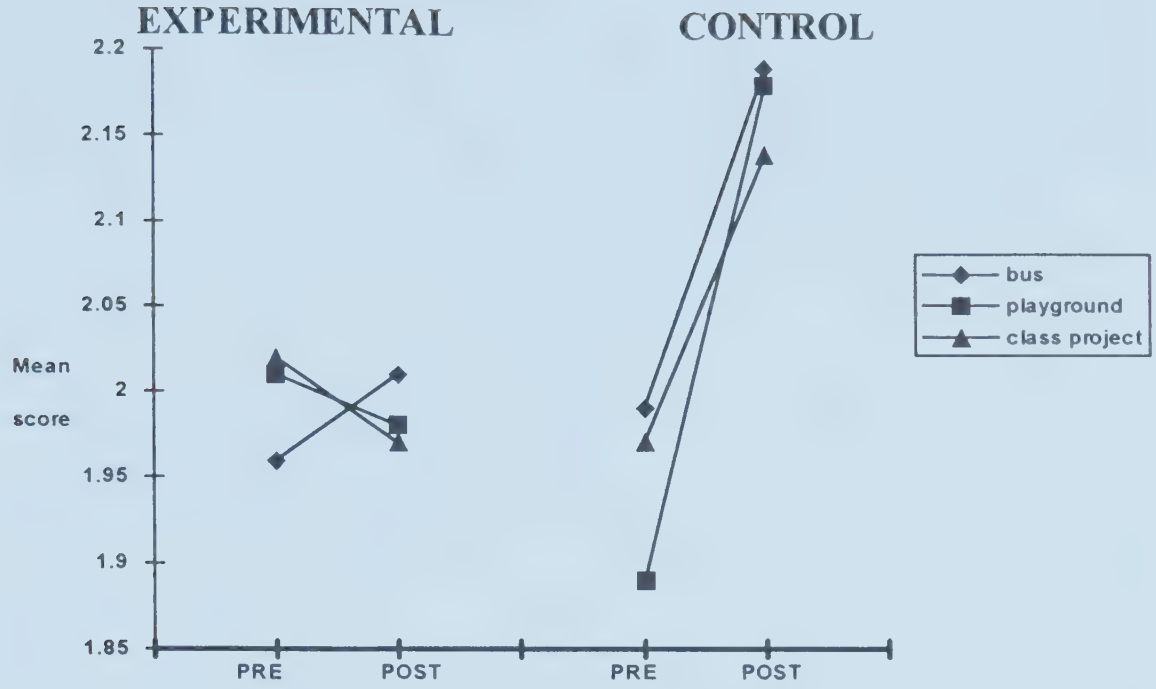


Figure 2. Social acceptance as a function of treatment conditions.

TEST SESSION: with 2 levels: pretest and posttest). The results of this analysis indicated that there was a significant main effect of group ($F(1,25)=13.60$, $p.<01$) indicating lower sociometric status in the control group and a group x test session interaction ($F(1,25)=7.47$, $p.<02$). Decomposition of this effect indicated that there was no change in the experimental group from pre- to posttest ($F(1,11)=.11$, $p>.5$), *but there was a significant decline in status in the control group* ($F(1,14)=19.63$, $p<.001$).

It was hypothesized that the beneficial effects of the program on social acceptance would be mediated by improvements in social problem solving strategies. If this hypothesis were true, then one would predict significant correlations between social problem solving and the sociometric measures at posttest when controlling for the pretest correlation between these two variables. To examine this issue, a partial correlational analysis was conducted on the social problem solving data for the two situations for which there were significant treatment effects (name calling and classmates laughing) and the sociometric "playground" measure. The results of this analysis indicated that from pre- to posttest there was greater stability for the sociometric measure ($r(25)=.87$) than for the problem solving measure (*R's* (25)=.25 and .29 for the name calling and peer laughing situations respectively). Next, for both groups combined there were no significant correlations between responses to the hypothetical situations and sociometric status at pretest for either situation (*R's* (25)=.08 and .01 for the name calling and peers laughing situations respectively). At posttest, these correlations increased to .27 and .24. The partial correlations were .30 and .21, neither of which is significant, suggesting that from pre- to posttest there was not a significant increase in variance in the sociometric measure accounted for by the social problem solving measure. Analysis of the separate groups indicated that in the name calling situation both experimental and control subjects showed an increase in the magnitude of the correlation

from pre- to posttest, but in neither case was that increase significant (partial $r(25)=-.35$ for the experimental group and .30 for the control group). For the peers laughing situation in the goal-given condition the correlation was stronger for the control group at pre- and posttest, but only the experimental group showed a significant increase in the value of the correlation from pre- to posttest (as indicated by the partial $r=.42$, $p<.05$). Thus for the one situation (peers laughing) it appeared that there was a program-induced increase in the relationship between problem solving strategies and sociometric measures that suggests generalization of the treatment effects.

Discussion

The results of the present study indicate that a social skills intervention program with emphasis on coaching, role playing, and information sharing conducted in an intact classroom can have some positive impact on the social acceptance and social problem solving skills of students with learning disabilities. In contrast, previous studies that have targeted students with learning disabilities for intervention and that have led to successful outcomes (see McIntosh et al., 1991) have been conducted in groups pulled out of their classrooms for intervention.

In terms of specific hypotheses, it was proposed that the intervention would have positive and significant effects on social acceptance. This hypothesis was not confirmed insofar as there was no positive change in the sociometric ratings in the experimental group. However, there was some evidence that the program was effective because there was a significant decline in sociometric ratings from pre- to posttest in the control group, but not the experimental group. This result may be a manifestation of a previous finding by Bierman et al. (1987) of a decline in the number of positive peer interactions over time in rejected boys who do not receive treatment. It may be that because there was no formal social skills program in the control group, the overall level of peer acceptance may have declined over the course of the year. A social skills program of the type employed in the present study may prevent such a deterioration in peer relationships. To verify this possibility, future studies should examine the relationship between the number of positive peer interactions and sociometric status in children with learning disabilities who are and who are not receiving social skills training.

Analysis of the data from the hypothetical situations indicated two situations that were especially problematic for the students in both groups: name calling and being laughed at by other students when performing poorly at a game. This result is consistent with evidence presented by Dodge, Petit, McClaskey, and Brown (1986) and Lochman and Lampron (1986) that situations involving provocation pose the greatest difficulty for aggressive children. The present study extends this finding for children with learning disabilities. In the present study these two situations were the only ones in which there were significant treatment effects: Children in the experimental group evidenced a significant improvement in the quality of problem solving strategies used at post- as compared to pretest in both of these situations. This effect suggests that the intervention was effective in improving the social problem solving skills of children in the experimental group, whereas no such effect was obtained in the control group. Carlson (1987) found that male children with learning dis-

abilities were more likely to select negative strategies than positive strategies to resolve conflict. Insofar as all the hypothetical situations used in the present study were conflictual in nature, the present study extends this finding by showing that intervention can increase use of positive strategies in conflict situations with children who have learning disabilities. It might be argued that the treatment effects on the hypothetical situations were weak because significant results were obtained from only two out of six situations. However, reference to Figure 1 shows that prior to treatment students used significantly lower-level strategies in the two situations where there were significant treatment effects. It seems reasonable to suppose that treatment effects may not have been realized in the other four situations because of ceiling effects.

It was hypothesized that the change in sociometric status would be mediated by improvement in social problem solving strategies. In the present study one significant relationship emerged; that is, when the pretest correlation between the sociometric measure and the social problem solving measure was removed by means of a partial correlation, there was a significant relationship between sociometric status and social problem solving for one situation only (peers laughing at the student). Thus there is some evidence in the present study that improvements in peer relationships were mediated by improvements in social problem solving skill. Because the relationship was quite weak and restricted to one situation, it is important to replicate this finding with another sample. The need for replication is underscored in this situation because in previous studies it has been the rule that social problem solving is not significantly related to measures of adjustment derived from teacher and/or peer perceptions (Dodge et al., 1986; Rickel, Eshelman, & Loigman, 1983; Sharp, 1981; Vitaro & Pelletier, 1991; Weissberg et al., 1981). The one exception is the study conducted by Gottman, Gonso, and Rasmussen (1975) in which a positive relationship was obtained between the sociometric measure and measures of both referential communication and knowledge of how to make friends.

Although the present study provides limited support for the hypothesis that a coaching/role playing/information sharing intervention can have impact on peer acceptance, the mechanism that mediates this effect is not clear. Although there was some evidence that the effects of the program on peer acceptance may have been mediated by changes in social problem skills, this effect was quite weak. The small magnitude of this effect may have been because cognitive and metacognitive development in students is multifaceted, requiring substantial time and effort in students and teachers. The only explicitly cognitive component of the intervention occurred during the debriefing sessions following the role-plays where students were encouraged to generate strategies, stop and think before they act, and monitor and evaluate their strategies. It may be that the connection between cognitive and metacognitive development and peer acceptance could be strengthened in future studies by targeting a greater variety of dimensions of cognitive and metacognitive strategies and by providing more depth in teacher training and intervention. It is also likely that in order to accurately measure the impact of a social skills intervention, it may be necessary to evaluate skill acquisition in real rather than hypothetical situations. This proposal is consistent with Vitaro and Pelletier's

(1991) finding that indicated weak relationships between performance in hypothetical and real social situations.

It is also possible that the magnitude of the treatment effects was limited by the fact that there may have been insufficient attention paid to student attributions. In the present study, the focus of attention was on strategy use and metacognitive awareness, but less so on attribution. Prior studies have suggested that many children with learning disabilities have motivational problems that stem from failure experiences and a tendency to attribute these failures to insufficient ability (Butkowski & Willows, 1980; Licht, 1983). Treatment programs that teach students to attribute past failures to insufficient effort tend to foster effort attributions and increased task persistence (Andrews & Debus, 1978; Dweck, 1975). Others have shown that effort feedback related to successful achievement enhances an individual's motivation and skills (Schunk, 1982). Borkowski and his colleagues have investigated the relationship between cognitive and motivational processes by examining the effects of attributional retraining and strategy training on children with learning disabilities (Borkowski, Weighing, & Carr, 1988). Their results showed that strategy training alone, without correcting negative attributional beliefs, is an ineffective method of instructing students with learning disabilities and is not likely to promote treatment generalization (Borkowski, Estrada, Milstead, & Hale, 1989). This finding underscores the importance of integrating metacognitive instruction and attributional retraining in programs designed to improve the generalized problem solving ability of children with learning disabilities. Future research should examine the impact of the executive functioning and attributional beliefs on the strategic behavior, generalized problem solving skills, and sociometric status of children with learning disabilities.

Training students with learning disabilities to become more effective and efficient social problem solvers is a multifaceted and challenging task that requires attention to cognitive, affective, and behavioral dimensions. Students need to be taught specific strategies and when and where to use them. Teachers need to facilitate the ability of students to monitor their own performance, develop skills in modifying and evaluating their strategic behavior, and provide them with feedback and practice. Moreover, intervention programs should be integrated in the school curriculum and include skills for children with learning disabilities that when applied will result in beneficial social outcomes such as enhanced feelings of self-worth and improved social status.

To date little research has been directed at investigating the relationship between the perceptions of social situations and strategy use in children with learning disabilities. Hence more research is needed to specify those social situations that children find difficult, why they find these situations difficult, and how this affects their social problem solving skills.

The results of this study provide some support for the utility of social interventions to ameliorate the social difficulties of children with learning disabilities, and extend the results of previous pull-out interventions to the whole classroom. However, comprehensive conceptual models that delineate the mechanisms by which social interventions affect the patterns of social acceptance of children with learning disabilities need further elaboration (Ladd & Mize, 1983; Pearl, 1987). Future research should examine the differential

effects of program components such as coaching, role playing, and information sharing on the social development of children with learning disabilities. In the present study, although some significant effects were obtained with the treatment, we are unable to discuss the relative contribution of the components of the program.

A number of methodological shortcomings to this study limit the conclusions to be made. First, because intact classrooms were used, it was not possible to randomly assign students to treatment conditions. Thus it is possible that the results were due to sampling bias. The only evidence against sampling bias is that on most measures that were taken (i.e., sociometric measures, social problem solving, academic functioning) there were no pretest differences. A second limitation of this study is that the control group was approximately one year older than the experimental group. To control for any effects of age, analyses of covariance were used whenever possible.

Although the results of this study must be considered tentative due to the above shortcomings, the study does expand on prior research on social intervention with children who have learning disabilities. One of the conclusions from this study is that the situation-specific social problem solving difficulties of children with learning disabilities should be more comprehensively delineated to better understand their social behaviors. This might allow for the design of treatments that can be individually tailored to meet the unique social needs of each child. This is consistent with the profile approach for identifying social skills and deficits proposed by Dodge and Murphy (1984) and the intervention approach of fitting social skills programs to target groups proposed by Coie (1985). Moreover, the present findings point to the need for more closely evaluating the relationship between perception of social difficulty and strategy use.

In the present program of research, it was evident from informal observations that limitations in children's expressive language skills influenced their ability to convey their ideas in coaching and role playing situations and to share information with their teachers and peers. Although expressive and receptive language skills were not included as variables in the research design, there is a need for future studies to evaluate social interventions with respect to children's communication skills across a range of social situations. This is consistent with Bryan's (1991) recommendation for research to link communicative and social competence to sociometric status.

This investigation did not evaluate whether the components of the social intervention program (coaching, role playing, and information sharing) were utilized by the teachers beyond the two days a week over the six-month period. Future studies should explore the generalization of the social learning process to the academic learning process.

References

Andrews, A., & Debus, R. (1978). Persistence and causal perception of failure: Modifying cognitive attributions. *Journal of Educational Psychology, 70,* 154-166.

Andrews, J. (1989). Social problems of learning disabled children: A review of the literature and some suggestions for assessment, intervention and future research. In C. Violato & A. Marini (Eds.), *Child development: Readings for teachers* (pp. 271-293). Calgary, AB: Detselig.

Association for Children with Learning Disabilities. (1985). ACLD description: Specific learning disabilities. *ACLD Newsbriefs,* 15-16.

Baum, D.D., Duffelmeyer, F., & Geelan, M. (1988). Resource teachers' perceptions of the prevalence of social dysfunction among students with learning disabilities. *Journal of Learning Disabilities, 21*, 380-381.

Bierman, K.L. (1986). Process of change during social skills training with pre-adolescents and its relation to treatment outcome. *Child Development, 57*, 230-240.

Bierman, K.L., & Furman, W. (1984). The effects of social skills training and peer involvement on the social adjustment of pre-adolescents. *Child Development, 55*, 151-162.

Bierman, K.L., Miller, C., & Stabb, S. (1987). Improving the social behavior and peer acceptance of rejected boys: Effects of social skill training with instructions and prohibitions. *Journal of Consulting and Clinical Psychology, 55*, 194-200.

Borkowski, J.G., Estrada, M.T., Milstead, M., & Hale, C.A. (1989). General problem solving skills: Relations between metacognition and strategic processing. *Learning Disability Quarterly, 12*, 57-70.

Borkowski, J.G., Weighing, R.S., & Carr, M. (1988). Effects of attributional retraining on strategy based reading comprehension in learning disabled students. *Journal of Educational Psychology, 80*, 46-53.

Bos, C., & Filips, D. (1982). Comprehension monitoring skills in learning disabled and average students. *Topics in Learning and Learning Disabilities, 2*, 79-85.

Bruck, M., & Hebert, M. (1982). Correlates of learning disabled students' peer interaction patterns. *Learning Disability Quarterly, 5*, 353-362.

Bryan, T. (1991). Social problems and learning disabilities. In B.Y.L. Wong (Ed.), *Learning about learning disabilities* (pp. 195-229). New York: Academic Press.

Burns, P.C., & Roe, B.D. (1985). *Burns-Roe informal reading inventory: Pre-primer to grade 12*. Boston, MA: Houghton Mifflin.

Butkowsky, I.S., & Willows, D.M. (1980). Cognitive-motivational characteristics of children varying in reading ability: Evidence of learned helplessness in poor readers. *Journal of Educational Psychology, 72*, 408-422.

Carlson, C.I. (1987). Social interaction goals and strategies of children with learning disabilities. *Journal of Learning Disabilities, 20*, 306-311.

Coie, J.D. (1985). Fitting social skills intervention to the target group. In B.H. Schneider, K.H. Ruben, & J.E. Ledingham (Eds.), *Children's peer relations: Issues in assessment and intervention* (pp. 141-156). New York: Springer-Verlag.

Conners, C.K. (1969). A teacher rating scale for use in drug studies with children. *American Journal of Psychiatry, 126*, 884-888.

Conte, R., & Andrews, J. (1993). Social skills in the context of learning disability definitions: A reply to Gresham and Elliott and future directions. *Journal of Learning Disabilities. 26*, 146-153

Dodge, K.A., & Murphy, R.R. (1984). The assessment of social competence in adolescents. In P. Karoly & J.J. Steffan (Eds.), *Adolescent behavior disorders: Current perspectives* (pp. 61-96). Lexington, MA: Heath.

Dodge, K.A., McClaskey, C.L. &, Feldman, E. (1985). Situational approach to the assessment of social competence in children. *Journal of Consulting and Clinical Psychology, 53*, 344-353.

Dodge, K.A., Petit, G.S., McClaskey, C.L. &, Brown, M.M. (1986). Social competence in children. *Monographs of the Society for Research in Child Development, 51*, 1-85.

Dudley-Marling, C.C., & Edmiaston, R. (1985). Social status of learning disabled children and adolescents. A review. *Learning Disability Quarterly, 8*, 189-204.

Dweck, C.S. (1975). The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology, 31*, 674-685.

Forness, S.R., & Kavale, K.A. (1991). Social skills deficits as primary learning disabilities: A note on problems with the ICLD diagnostic criteria. *Learning Disabilities Research and Practice, 6*, 44-49.

Fox, C. (1989). Peer acceptance of learning disabled children in the regular classroom. *Exceptional Children, 56*, 50-57.

Gorney-Krupshaw, B., Atwater, J., Powell, L., & Morris, E.K. (1981). *Improving social interactions between learning disabled adolescents and teachers: A child effects approach* (Research Report No. 45). Lawrence, KA: University of Kansas Institute for Research in Learning Disabilities.

Gottman, J., Gonso, J., & Rasmussen, B. (1975). Social interaction, social competence, and friendship in children. *Child Development, 46*, 709-718.

Gresham, F.M. (1988). Social competence and motivational characteristics of learning disabled students. In M.C. Wang, M.C. Reynolds, & H.J. Walberg (Eds.), *Handbook of special education:*

Research and practice, Vol. 2, Mildly handicapped conditions (pp. 283-300). New York: Pergamon Press.

Gresham, F.M., & Elliott, S.N. (1989). Social skills deficits as a primary learning disability. *Journal of Learning Disabilities, 22*, 120-124.

Gresham, F.M., & Nagle, R.J. (1980). Social skills training with children: Responsiveness to modeling and coaching as a function of peer orientation. *Journal of Consulting and Clinical Psychology, 18*, 718-729.

Gresham, F.M., & Reschley, D.J. (1986). Social skill deficits and low peer acceptance of mainstreamed learning disabled children. *Learning Disability Quarterly, 9*, 23-32.

Hallahan, D.P., & Bryan, T.H. (1981). Learning disabilities. In J.M. Kaufman and D.P. Hallahan, (Eds.), *Handbook of special education* (pp. 122-153). Englewood Cliffs, NJ: Prentice-Hall.

Hallahan, D.P., & Reeve, R.E. (1980). Selective attention and distractibility. In B.K. Keogh (Ed.), *Advances in special education, Vol. 1: Basic constructs and theoretical orientations* (pp. 141-181). Greenwich, CT: JAI Press

Havertape, J.F., & Kass, C.F. (1978). Examination of problem solving in learning disabled adolescents through verbalized self-instructions. *Learning Disability Quarterly, 1*, 94-99.

Idol, L., Paolucci-Whitcomb, P., & Nevin, A. (1986). *Collaborative consultation*. Austin, TX: Pro Ed.

Interagency Committee on Learning Disabilities. (1987). *Learning disabilities: A report to the U.S. Congress*. Washington DC: U.S. Department of Health and Human Services.

Johns, J. (1985). *Basic reading inventory: Pre-primer to grade eight*. Dubuque, IA: Kendal Hunt.

Ladd, G.W. (1981). Effectiveness of a social learning method for enhancing social interactions and peer acceptance. *Child Development, 52*, 171-178.

Ladd, G.W., & Mize, J. (1983). A cognitive-social learning model of social skills training. *Psychological Review, 90*, 127-157.

Lambert, N., & Sandoval, J. (1980). The prevalence of learning disabilities in a sample of children considered hyperactive. *Journal of Abnormal Child Psychology, 8*, 33-50.

Licht, B.G. (1983). Cognitive-motivational factors that contribute to the achievement of learning disabled children. *Journal of Learning Disabilities, 16*, 483-490.

Lochman, J.E., & Lampron, L.B. (1986). Situational social problem solving skills and self-esteem of aggressive and non-aggressive boys. *Journal of Abnormal Child Psychology, 14*, 605-617.

McIntosh, R., Vaughn, S., & Zaragoze, N. (1991). A review of social interventions for students with learning disabilities. *Journal of Learning Disabilities, 8*, 451-458.

Oden, S., & Asher, F.R. (1977). Coaching children in social skills for friendship making. *Child Development, 48*, 495-506.

Pearl, R. (1987). Social cognitive factors in learning disabled children's social problems. In S.J. Ceci (Ed.), *Handbook of cognitive, social, and neurological aspects of learning disabilities* (vol. 2, pp. 273-299). Hillsdale, NJ: Erlbaum.

Pressley, M., & Levin, J.R. (1987). Elaborative learning strategies for the inefficient learner. In S.J. Ceci, (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities* (vol. 1, pp. 175-212). Hillsdale, NJ: Erlbaum.

Rickel, A.U., Eshelman, A.K., & Loigman, G.A. (1983). Social problem solving training: A follow-up study of cognitive and behavioral effects. *Journal of Abnormal Child Psychology, 11*, 15-28.

Schneider, B.H. (1992). Didactic methods for enhancing children's peer relations: A quantitative review. *Clinical Psychology Review, 12*, 363-382.

Schumaker, J.B., & Ellis, E.S. (1982). Social skills training of LD adolescents: A generalization study. *Learning Disability Quarterly, 5*, 409-414.

Schumaker, J.B., & Hazel, J.S. (1984). Social skills assessment and training for the learning disabled: Who's on first and what's on second. *Journal of Learning Disabilities, 17*, 422-430.

Schunk, D.H. (1982). Effects of attributional feedback on children's perceived self-efficacy and achievement. *Journal of Educational Psychology, 74*, 548-556.

Sharp, K.C. (1981). Impact of interpersonal problem-solving training on pre-schoolers' social competence. *Journal of Applied Developmental Psychology, 2*, 129-143.

Sheinker, J., & Sheinker, A. (1988). *A metacognitive approach to social skills training*. Rockville, MD: Aspen Publications.

Torgesen, J.K. (1980). Conceptual and educational implications of the use of efficient task strategies by learning disabled children. *Journal of Learning Disabilities, 13*, 19-26.

Vaughn, S. (1985). Why teach social skills to learning disabled students? *Journal of Learning Disabilities, 18*, 588-591.

Vaughn, S., & La Greca, A.M. (1988). Social skills of LD students: Characteristics, behaviors and guidelines for intervention. In K. Kavale (Ed.), *Handbook in learning disabilities* (pp. 123-140). San Diego, CA: College Hills.

Vitaro, F., & Pelletier, D. (1991). Assessment of children's social problem solving skills in hypothetical and actual conflict situations. *Journal of Abnormal Child Psychology, 19*, 505-518.

Weissberg, R.P., Gesten, E.L., Rapkin, B.D., Cowen, E.L., Davidson, E., Flores deApodaca, R., & McKim, B. (1981). Evaluation of a social-problem-solving training program for suburban and inner-city third-grade children. *Journal of Consulting and Clinical Psychology, 49*, 251-261.

West, J.F., & Cannon, G.S. (1988). Essential collaborative consultation competencies for regular and special educators. *Journal of Learning Disabilities, 21*, 56-63.

West, J.F., & Idol, L. (1990). Collaborative consultation in the education of mildly handicapped and at-risk students. *Remedial and Special Education, 11*, 22-31.

West, J.F., Idol, L., & Cannon, G. (1989). *Collaboration in the schools: An inservice and preservice curriculum for teachers, support staff, and administrators*. Austin, TX: Pro-Ed.

Whang, P. L., Fawcett, S.B., & Matthews, R. (1984). Teaching job-related skills to learning disabled adolescents. *Analysis and Intervention in Developmental Disabilities, 4*, 29-38.

Wiener, J. (1987). Peer status of learning disabled children and adolescents: A review of the literature. *Learning Disabilities Research, 2*, 62-79.

Wong, B.Y.L., Wong, R., Perry, N., & Sawatsky, D. (1986). The efficiency of a self-questioning summarization strategy for use by underachievers and learning disabled adolescents. *Learning Disability Focus, 2*, 20-35.

The Alberta Journal of Educational Research *Vol. XLI, No. 1, March 1995, 103-121*

John A. Ross
Ontario Institute for Studies in Education

and

J. Bradley Cousins
University of Ottawa

Giving and Receiving Explanations in Cooperative Learning Groups[1]

This qualitative study examined patterns of giving and receiving explanations in three samples of students working in cooperative learning groups (N=65 grade 7/8, 45 grade 9/10, and 51 grade 9/10) on problem solving tasks (correlational reasoning problems). Students were randomly assigned to mixed ability cross-sex groups of four or pairs; the treatment was Slavin's STAD. A gap was observed between student practice and an ideal model of explanation exchanges. Students infrequently sought explanations. They gave explanations that were inadequate because they were uncertain about performance, lacked a language for describing their thoughts, had deficient teaching skills, and were insensitive to the needs of help seekers. Explanation givers did not monitor recipients' understanding of explanations. The article proposes that the benefits of cooperative learning will be enhanced if teachers teach students how to ask for and give quality explanations.

Cette étude qualitative a examiné les tendances qui ressortaient lorsqu'on donne ou lorsqu'on reçoit des explications en se basant sur trois échantillons d'élèves qui travaillaient en groupe dans le contexte de l'apprentissage coopératif en groupes (N=65 $7^e/8^e$ année, N=45 $9^e/10^e$ année, et N=51 $9^e/10^e$ année) sur des tâches de résolutions de problèmes (raisonnement de problèmes corrélationnels). En pigeant les élèves au hasard, on les sépara en paires ou en groupes de quatre ayant des habiletés mixtes et des deux sexes à la mode de Slavin selon le traitement STAD (Student-Teams-Achievement-Divisions), c'est-à-dire, en regroupant les élèves en équipes qui visent à l'atteinte d'objectifs particuliers. On observa un écart entre la pratique actuelle des élèves et un modèle idéal des échanges des explications possibles. Les élèves recherchaient rarement des explications. Ils donnaient des explications qui étaient inadéquates parce qu'ils étaient incertains de leur performance, n'avaient le langage approprié pour décrire leurs pensées, étaient déficients en habiletés pédagogiques, et étaient insensibles aux besoins de ceux qui sollicitaient de l'aide. Les élèves qui donnaient des explications ne surveillaient pas le niveau de compréhension de ces explications de ceux qui les recevaient. Cet article propose que les bénifices de la démarche coopérative seraient accrus si les enseignant(e)s enseignaient aux élèves à demander des explications et à et donner des explications de qualité.

The constructivist revolution in instruction (Posner, Strike, Hewson, & Gertzog, 1982) has stimulated interest in teaching methods in which students learn from one another. This article reports a qualitative investigation that

John Ross is a professor of curriculum and Head of the OISE Trent Valley Centre. His research interests are curriculum development, program evaluation, and student assessment.
J. Bradley Cousins is an associate professor. His research interests are knowledge utilization, organizational learning, and evaluation.

examined the explanations that students give and receive when they are working together to learn a new skill.

Theoretical Framework

The Importance of Explanations

Process-product studies of student-student interactions (the most complete reviews are in Webb, 1989, 1992) have consistently found that students who give explanations learn more than those who do not, even after student ability is controlled. Learning occurs because giving explanations requires the reorganization of the material to be learned (Bargh & Schul, 1980); it makes unconscious thoughts explicit (King, 1989, 1990), and it contributes to conceptual conflict that can lead to cognitive restructuring (Doise & Mugny, 1979). In contrast, receiving an explanation has a variable impact, regardless of whether or not an explanation is received in response to a specific request. Webb (1989) observed:

> even receiving help at what appears to be a sufficiently high level of elaboration is not often sufficient for high achievement.... Student interaction would have to be examined more closely to determine whether the explanations were relevant, understood, and applied and internalized by the target student. (p. 28)

Webb's suggestion focuses attention on the adequacy of explanations that are given. The low impact of explanations might also be related to the failure of help givers to oversee the recipients' use of explanations. Hooper (1992) found that the achievement of those who were given help was higher when the help givers checked recipients' understanding, but other studies have found that students do not often monitor each other's performance (Azmitia, 1988; Ross, 1994).

Another reason for the unproductive outcomes of receiving an explanation might be that students do not ask for explanations when it is appropriate to do so. Students do not know when they do not understand (Markman, 1977, 1979) and do not seek help when it is needed (Good, Slavings, Harel, & Emerson, 1987; Karenbenick & Knapp, 1988; Newman, 1990). Requests for explanation are infrequent in most studies. The highest means were reported by Webb and Kenderski (1985). They found that a grade 8-9 subsample of high ability girls averaged nine requests for explanation per 45-minute period. The means reported by other studies, and for the remaining students in Webb and Kenderski (1985), were dramatically lower. For example, Hertz-Lazarowitz (1989) found that students in grades 3-8 engaged in high-level discussions less than 2% of the time; Deering and Meloth (1990) found that 91% of academic help interactions were at an information exchange level. Kempa and Ayob (1991) found that only 10% of information exchanges of able 16-year-olds were at the explainer level; and Corno (1989) observed that fewer than 5% of the interactions of 5th graders involved instructing.

It is also possible that students do not know how to ask for explanations. Several studies have reported that characteristics of the question (such as being repeated, specific, and directed toward a particular individual) determine whether it will be answered (Wilkinson & Calculator, 1982; Wilkinson & Spinelli, 1983).

These findings suggest that receiving an explanation is a productive form of cognitive engagement only if three conditions are met. First, explanation seekers must recognize a need for help, diagnose its specific dimensions, locate a suitable help giver, and access the peer resource through an appropriate request. Second, the explanation seeker must receive a quality explanation, consisting, for example, of the peer modeling the correct performance while providing an oral commentary that describes how the performance was generated, using language and ideas familiar to the learner. The specific teaching moves of effective explanation givers may vary, approximating techniques such as participant modeling instruction (Corno, 1992), cognitive apprenticeship (Brown, Collins, & Diguid, 1989), and expert scaffolding (Palincsar & Brown, 1984). A quality explanation is one that gives the specific answer required by the immediate task while providing the student with sufficient procedural knowledge to solve problems of a similar type on his or her own in the future. Third, the explanation seeker needs to internalize the explanation by applying it to the learning task and obtaining corrective feedback.

Cooperative Learning

Cooperative learning approaches differ in the attention they give to training students in group interaction skills. For example, Johnson and Johnson (1987) suggest that practice and reinforcement of peer tutoring skills are essential. Their approach provides an extensive array of activities to identify, demonstrate, and consolidate skills prerequisite to students learning from one another (Johnson, 1990). Feedback on cooperative skills is also a critical component of group investigation (Sharan & Shachar, 1988) and of the eclectic approach of Kagan (1988). There is evidence that prosocial behaviors increase in these approaches (Johnson, Johnson, & Stanne, 1986), although the unique contribution of the social skills component has not been disentangled from other treatment dimensions.

In contrast, the methods of the Slavin research group (Teams-Games-Tournaments, Student-Teams Achievement-Divisions [STAD], Team Accelerated Individualization, and Cooperative Integrated Reading and Composition) provide for no explicit instruction in group interaction skills, and formal student feedback focuses on cognitive achievement alone. Teachers are advised, in STAD for example, to praise teams that are working well, but specific peer interaction skills are not negotiated with students or otherwise highlighted (Slavin, 1990). Slavin's reviews of the literature persuade him that the combination of individual and group rewards through the fair allocation of extrinsic rewards is sufficient to promote learning (Slavin, 1983, 1987, 1990, 1992).

> Students may be motivated to engage in elaborated, cognitively involving explanations and discussions if the learning of their group-mates is made important by the provision of group rewards based on individual learning performances. (Slavin, 1987, p. 1166)

> The evidence supports a conclusion that group member helping on a group task and group member norms supporting performance are consequences of cooperative incentive structures. (Slavin, 1992, p. 153)

Research Questions

The specific purpose of the research reported here was to observe students working in STAD groups. We wanted to find out if the high-level helping predicted by Slavin would be visible. Because no previous study has examined explanation exchanges in STAD groups, we were initially guided by the three conditions for effective exchanges described above. We selected a qualitative approach because we believed that process-product research on explanation giving and receiving has been limited in four ways. First, the quantitative approach breaks students' speech into discrete utterances threatening the coherence of the whole (Solomon, 1992). Second, most of the research has been conducted in settings that gave little support for student helpfulness. For example, very few of these investigations used reward interdependence to encourage explanation givers to attend to the needs of students who needed help or instituted individual accountability to reduce social loafing. Third, the tasks assigned to students were rarely challenging: learning routine algorithms, for example, is unlikely to stimulate high-level cognitive exchanges. Fourth, little attention (with the exception of Hooper, 1992 and Webb, 1992) has been given to students' use of the explanations they receive.

Method

Sample

Productive conditions for studying explanation exchanges were provided in school districts that were responding to changes in Ontario curriculum policy that required that students acquire correlational reasoning skills in grades 7-10 (Ontario Ministry of Education, 1988). Previous research has found that correlations are the most challenging of the formal reasoning tasks (Lawson & Bealer, 1984; Yap & Yeany, 1988). Coincidentally the province was promoting the implementation of cooperative learning in the same grades as a mechanism for dealing with the elimination of ability tracking in the formation of classes (Ontario Ministry of Education, 1992).

There were three samples: (a) 65 grade 7 and 8 students working in groups of four in four classrooms using STAD; (b) 45 grade 9 and 10 students in three classes in a similar setting; and (c) 51 grade 9 and 10 students in three classes with each pair sharing a computer following a STAD approach. The average age was 13 years for sample (a) and 15 years for samples (b) and (c). Students were drawn from schools in four districts in central Ontario. Most schools were in small cities and towns. None was from a large city and there were few nonwhite students in the samples. Males and females were equally represented except in sample (b) which had slightly more females.

Instruments

Student interactions while in cooperative groups were audio- [samples (b) and (c)] or videotaped [sample (a)] on a single occasion at the end of the program for 25-45 minutes (the average was 30 minutes). The qualitative analysis of these data is reported in this article. In addition, quantitative data (student achievement and attitudes toward helpfulness) were collected. These data are not reported here; findings are described in Ross and Cousins (in press).

Treatment

In designing the treatments STAD procedures were followed closely. Materials were prepared using direct instruction modules that produced high levels of achievement in earlier experiments (Cousins & Ross, 1993; Ross & Cousins, 1993a, 1993b). A procedure for solving two variable and multivariate correlational problems was demonstrated by the teacher. Students were given practice activities that required the construction of graphs using content from the grade 7/8 or grade 9/10 geography curriculum. Students were assigned to teams as detailed in Slavin (1990): Teachers rank ordered students from least to most capable, using students' past performance on tests and observations of student behavior. The list was divided into ability quartiles and one student was randomly assigned to each group from each quartile. The resulting teams were adjusted by teachers to ensure a balance of genders and ethnicities and to avoid delinquent alliances (replacements were made within ability quartiles). Students in samples (a) and (b) completed practice activities in mixed-sex, mixed-ability groups of four. Students in sample (c) completed the same activities in mixed-sex, mixed-ability pairs; each pair shared a computer, using the software program CORReoGRAPH.[2]

In all treatments students worked in cooperative groups for 70% of the 300-minute duration of the program. During each practice period students were given two exercises to work on. The student task consisted of constructing scatterplots and trend lines to represent the relationship between two continuous variables, in some instances while controlling for a third variable. Figure 1 provides an example of a correlational reasoning task; Figure 2 provides an example (produced by CORReoGRAPH) of an answer to the exercise. Students in samples (a) and (b) completed all activities by hand; students in sample (c) did a few examples by hand, but most were created by students on the computer.

Each student was expected to complete his or her own graphs, asking for and receiving help when it was required. Near the end of each period teachers distributed answers to the exercises. Students were told that the purpose of the group activities was to learn how to solve correlational problems that would be assessed in two quizzes. Individuals completed each quiz without assistance from team members. Reward structures combined individual and group accountability using Slavin's (1990) fair sharing procedure that makes individual performance visible and combines each group member's improvement scores in calculating group rewards. The rewards varied between classes but all were extrinsic, most frequently consisting of team certificates, stickers and extra computer time. Students stayed in the same groups for the duration of the correlational reasoning module.

The treatments differed in one significant way from the original STAD. Although students had been working in groups previously, they had not been placed in structured STAD teams. Consequently the duration of the treatment was less than what is recommended for full STAD implementation.

Procedure

Verbatim transcripts of the audio/videotapes were made and read in their entirety by members of the research team. Attention focused on the explanation exchange protocols (which included interactions in which explanations

People Make Garbage

Angel and Tamil both live near a landfill site and they are very concerned about the amount of garbage that is produced in their community.

Angel: *Some people just don't care. The family next door to us puts out at least 8 bags of garbage every week. There ought to be a law that nobody can put out more than 3 garbage bags per week.*

Tamil: *Not so fast, Angel. What matters is the weight of garbage, not how many bags it is in. And what about the size of the family? Large families probably produce more garbage than small families.*

Angel: *I don't think that's true about the size of the family. What counts is whether they are recycling.*

Angel and Tamil decided to investigate the question. They collected information on the weight of garbage put out by each household. They recorded the number of people in the household and recycling habits (whether or not the household used a blue box and a compost pile).

Use the information collected to find out if there is a relationship between size of household and amount of garbage produced.

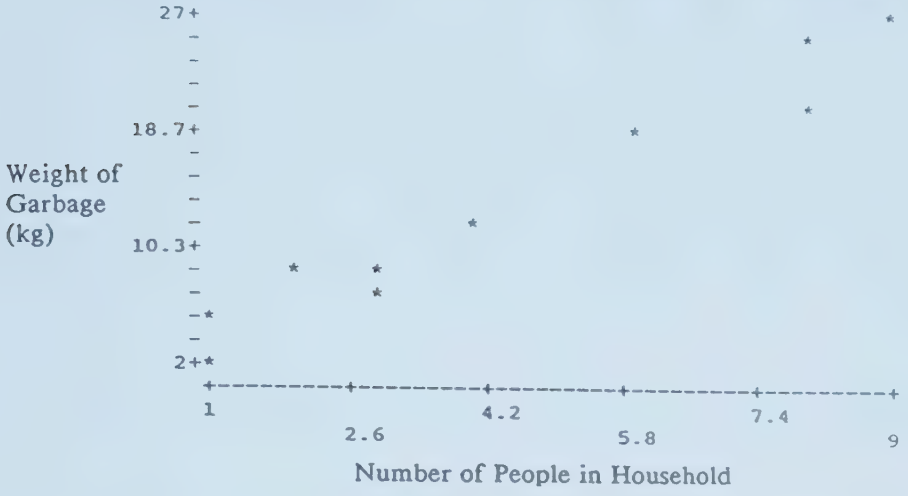
Household	*Number of Members*	*Weight of Garbage (Kg)*	*Recycling Habits*
Armitage	2	8	no
Balson	6	18	no
Carlson	3	2	yes
Duncan	9	27	no
English	2	3	yes
Fuller	1	2	no
Gillespie	4	2	yes
Horvath	7	6	yes
Ibey	10	8	yes
Johnstone	1	5	no
Kay	3	3	yes
Lawler	8	5	no
Munson	2	6	yes
Newman	3	1	yes
O'Toole	4	12	no
Porter	2	2	yes
Roberts	4	4	yes
Scott	3	7	no
Thompson	3	8	no
Urbach	8	21	no

Figure 1. Example of correlational reasoning task.

were offered without an overt request). The request-explanation-practice model described above provided the general framework, and specific issues evolved as we made multiple passes through the data. The final set focused on the clarity, precision, and courtesy of the request, the requester's persistence, and his or her selection of a potential help giver. In inspecting the explanations that were offered, we focused on the fit between the explanation and the specific need of the request, the degree of elaboration, the suitability of examples that were used, the extent to which the explanation giver diagnosed the needs of the help seeker, the number of students involved in giving the ex-

GRAPH A

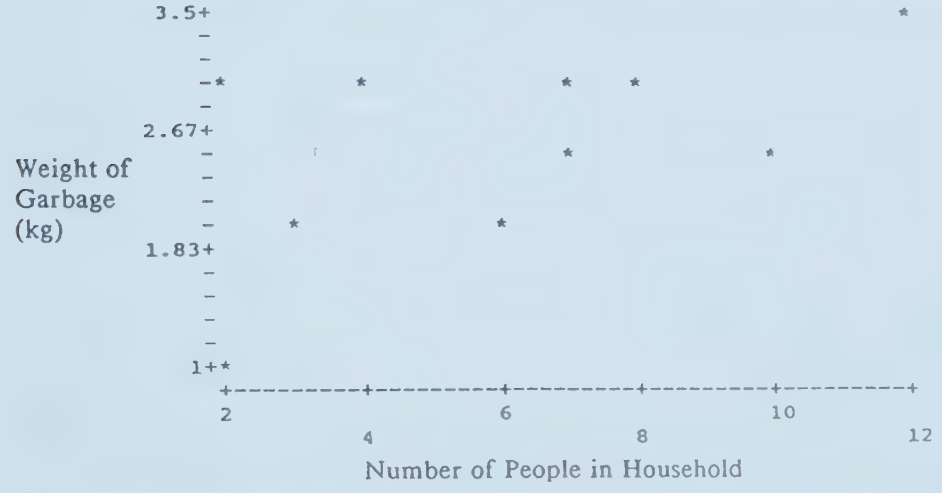
HOUSEHOLD SIZE AND GARBAGE PRODUCED FOR NON-RECYCLERS



For non-recycling households, there is a positive correlation. As the number of people in the household increases, the amount of garbage increases.

GRAPH B

HOUSEHOLD SIZE AND GARBAGE PRODUCED FOR RECYCLERS



For recycling households, there is a positive correlation. As the number of people in the household increases, the amount of garbage increases, but not to the same degree as in non-recycling households.

Figure 2. Example of answer to correlational reasoning exercise.

planation, the accuracy of the explanation, and the affective climate of the interaction. In examining instances of monitoring we focused on whether help givers asked if the explanation was understood, any tests of the help seekers' grasp of it, the assignment of practice activities, and attempts to reexplain. Although the samples were examined separately, the interpretive framework was shared and the results were highly similar. The findings are reported for the three samples combined, except where differences require a separate discussion.

Results

Explanation seeking. Only a third of the students requested an explanation and only half of these students received one in response. The requests for explanation were all sincere; none represented unnecessary help seeking. But requests tended to be poorly articulated. Few were directed toward a particular individual, most were brief, and many were vague, for example, "What are you doing?" In contrast, a student who saw that his graph did not have the same shape as one drawn by another group member, posed a much more precise request: "Since this [the distribution of data points] is much more spread out than yours, since I went up by ones [in scaling the axes], would [the teacher] still consider this a positive correlation?" Help seekers tended to be persistent, paraphrasing requests that were unanswered, and returning to the issue if they were unsatisfied. Requests that were specific and repeated were more likely to receive a response. Students who were unpleasant to their peers were less likely to receive assistance, and requests to students who were not members of the same pair or group were more likely to be ignored.

Explanation giving. There were several examples of explanations that were effective and brief. For example, a student who asked her partner why he was plotting different numbers than she was was told: "I'm doing the yes graph, that's why," meaning that he was plotting the same relationship (between family size and amount of household garbage), but for a different value of the control variable (recycle, yes or no). Similarly, a student who asked a peer why she was scaling the temperature variable in her graph from what appeared to be high numbers to low numbers, instead of the recommended procedure of scaling from low to high, was told "'cause negative 12 is colder than negative 1." There were also good longer explanations. In the following passage, Rachel [pseudonyms are used throughout] wanted to know why her graph did not look like that of another group member, in particular why she had more scale intervals on the bottom axis.

Rachel: What does yours look like, how many do you need across?

Loughery: I'll need a different number than you.

Rachel: Why?

Loughery: Because you're using different numbers than me. Your degrees only range from 4 to 9; mine go from 9 to 18. [i.e., Rachel's intervals are 5 degrees wide; Loughery's are 9 degrees].

Some students gave explanations that combined an oral commentary with a nonverbal demonstration. For example,

Sean: I don't know how to do these things.

Wesley: Okay. So we're doing the no's right? [i.e., they are plotting the households which do not recycle their waste]. So circle all your no's.... Now we go, okay, 2 and 8 [i.e., a household with 2 members producing 8 kg of garbage per week] right? So's all you have to do is go [up the y axis to] your 2, you go up 2, across [on the x axis to] 8. Right about there. ding, ding, ding

Students sometimes referred to a past example when giving a demonstration.

Ryan: How do you do it?

Stephanie: I'm not sure. [Pause] You guys. For this next one, I do the same as I did for this one? Okay guys?

Brandy: Sure.

Stephanie: So number of members is on the vertical axis and it's going up by 1, 2, 3, 4. And the horizontal axis is by 2s.

Some student explanations were probably too brief to be helpful to the help seekers. When Christa asked why all the data for the three variables given in the exercise could not go on the same graph, Natalie's explanation was that "it would be inconclusive if we did it like that." There were also explanations that contained misconceptions that probably confused the recipients. In the following Nate tried to describe how to distinguish between a negative and positive correlation.

Nate: That's positive. You know how you can tell? This way. As the graph goes up, as these numbers go up, these numbers go up too. If it was going like this, as these numbers go up, these ones would go. You know, how like, two positive make a

Valerie: negative.

Nate: yes.

Valerie: Positive. Two positive make a negative.

Nate: yes.

Valerie: Two negatives make a positive ...

Nate: And a positive plus a negative makes a negative ... as you're going up this way, both of these are going up, so it's positive [correct definition of a positive trend]. If you are going down that way, as this number goes up, this one would get lower, on here [correct definition of negative trend]. See? So a negative plus a positive times a positive will make a negative.

Valerie: Oh.

About 30% of the explanations contained errors. Some misconceptions in the explanations could have arisen from the models given in the lessons. For example, Christa believed that correlational predictions always had to be in the form recommended in the lesson, "if the temperature increases, then the snowfall will decrease," and that expressions like that of Wanda, "if the temperature decreases, then the snowfall will increase," were incorrect. Several students believed that when solving a problem involving more than two variables, the one that was in categorical form (e.g., inland/coastal location) should be selected as the control variable rather than making the decision on the basis of conceptual analysis.

Almost half the explanations appeared to be insensitive to the requestor's level of understanding. Some students persisted with explanations based on examples provided by the teacher even when it was inappropriate to do so. In the episode below Rachel did not understand why the group needed to produce four graphs when she thought three would do. The passage is particularly interesting because it shows how two students (Wanda and Loughery) worked together to help her even after she indicated that she should have known how to do it. Their intervention was unsuccessful; Rachel insisted later in the session that only three graphs were required.

Rachel: How do you get 4 [graphs]?

Wanda: You know how we did the soya bean one where you had to do, where there was degree days and averages.... Well, we have to figure out which one's the deciding factor for this one ... [when this example did not persuade Rachel, Wanda tried another]. You know when you did the cooks you grouped them all together and

Rachel: I wasn't here.

Loughery: Yes you were; [the teacher] showed it up on the overhead.

Wanda: Yesterday.

Rachel: Okay. I wasn't paying attention.

Wanda: Well maybe if you paid attention this wouldn't happen.

Loughery: Rachel, do you understand how he did that?

Rachel: No I don't.

Wanda: Remember when he put it on the overhead and he put all the cooks together and there was no correlation?

Rachel: Listen, just tell me what I have to do.

Wanda: You have to understand it to know what to do [she then rehearsed the teacher's lesson].

Explanation givers often confused help seekers, and at times themselves, by using different contexts to explain the procedures and by moving quickly from one instance to another. For example, Jacob was uncertain about how to set the scale for the axes in his graph. Jackie tried to help by demonstrating the skill, but she selected a different axis than the one he was working on and did not explain how she selected the upper end of the scale. She then tried to recall the example from the lesson without success.

Jacob: Now what am I going to go [up] by? 200?

Jackie: How [do you] break [it]? I forgot. I think you're supposed to put another break [i.e., scale interval] in it right like this.

Jacob: Just put another break after that? Why, what have you got there?

Jackie: Put the break there because it goes up to 20 [which was the value of the highest case in the data set].

Jacob: 20, 21, and then it goes out there.

Jackie: So where do you put another break?

Jacob: You can just keep going.

Jackie: No. You know how [the teacher] puts a break way up here?

Jacob: No I didn't.

Jackie: On the board he did.

Jacob: No he didn't. He just went up to a 1000 and put a 1000 there.

Jackie: Is that what I should do?

Jacob: I don't know. I don't even know what I'm doing.

Jackie: I'm pretty sure that's what you do.

Some opportunities for learning were also missed by help givers. For example, in one of the exercises students were given data on elevation, temperature, and latitude of weather stations on three continents. The stations were in the same latitude range on one continent and varied substantially in the other two data sets. They were asked to find the relationship between elevation and

temperature in each continent and to explain any differences they might find. In two of the groups students asked what they should do with the latitude data; in each case they were advised by their peers to ignore it.

Nathaniel: This is one of those things that they say. You're given more information than you need. We don't need this [latitude data]. Because [it's] just temperature and elevation.

On another occasion a student noticed that the pattern of data points in his graph was not the same as the pattern in the graph of his partner (because they used different scales). His partner was not interested in exploring why there might be differences.

Jason: What are you doing with them down here? Here we are up here.

Sandy: Well, that doesn't matter.

Many of the requests for explanations were met with off task behavior or the other members of the group continued with their own agendas. In at least one instance a request was ignored so that the group could finish: it was quicker to do it themselves than to explain it to someone else.

Wesley: Why don't you let me do it?

Gina: (to another group member) Come on, get to work. Hurry up so we can say we are done.

Requests for explanation were often met with procedural directives that dealt with the how, but not the why. For example, Ruth could not understand why the upper end of the scale in her partner's graph was so unlike her own. The partner responded to her request by describing his procedure, without explaining that he put a different variable on the axis.

Ruth: Why is it so small? That's not the highest [case].

Byron: We're doing this, the bottom one, by 3s; it works.

Explanation monitoring. Explanation givers were rarely observed testing the effects of their explanations. All instances of monitoring occurred in the grade 9/10 samples ([b] and [c]). One of the few examples consisted of a student (Anna) giving a lengthy recapitulation of the steps she and her partner (Randy) went through to complete one problem as a way of explaining how to do items of a similar type. The recitation appeared to be insufficient, but when Anna demanded that Dale attempt his own account of what they did he was able to do so succinctly.

Dale: [after the explanation] I still don't understand.

Anna: We did it right. We did it right.

Dale: OK, I think I understand now.

Anna: [skeptically] Explain it.

Dale: I can't. [but then he does] We did these two [pointing to household size and energy use, the variables to be correlated] with the no's [for the households in which there was no recycling]

Anna: Yeah.

Dale: And we did these two again with the yes [i.e., they correlated the same variables for the households in which there was recycling]

Anna: Yes

Dale: Oh, okay. I understand.

There was little co-construction of explanations; help seekers rarely contributed. Some help givers imitated the Socratic questioning used by their teachers to elicit understanding. This technique was unsuccessful in several instances because the questioners were at a loss as to how to continue the discussion when they received an unexpected response. Instead of continuing with another probe, the explanation providers gave up.

In the grade 7/8 sample there was no monitoring of the effectiveness of the explanations that were given. We could not find a single instance of an explanation giver testing to determine whether the explanation had been understood and appropriately used.

Discussion

Examination of specific interactions indicated a wide gap between the ideal model of explanation giving/receiving and student practice. The study had three main findings.

Students Infrequently Sought Explanations

Our investigation confirmed that students infrequently seek explanations, possibly because they did not know they needed help and did not know whom to ask if they did. Even though the correct answers to each exercise were distributed to students at the end of each group activity, students had difficulty interpreting differences between their work and the official answers. Graphs with apparently different data plots could express the same relationship because students were allowed to make their own decisions about the size and shape of the groups. The result was that students did not have good information on the performance of themselves or their peers. If students did not know they needed help, they would be unlikely to seek it (Markman, 1977, 1979).

The length of the treatment was also a factor in depressing explanation seeking. Constructing graphs was time consuming, especially for samples (a) and (b) who did everything by hand. The result was that students did not have a lot of data on which to make judgments about the competence of their peers. It also meant that they did not have enough interactions with group members to determine who would be willing to give help in a nonthreatening way, an important consideration as fear of ridicule is a major impediment to help seeking (Newman & Goldin, 1990; Newman & Schwager, 1993).

Students Received Inadequate Explanations

Many of the explanations were strewn with misconceptions. A reason is that they were as likely to be given by low- as high-ability students; neither rank in class nor specific competence (pretest scores on a correlational reasoning instrument) predicted explanation giving. In contrast, previous research on explanations has found that the more able give and the less able receive. Webb and Kenderski (1984) found that students with higher ability relative to other group members were more likely to give explanations, and Webb (1989) identified six studies in which giving explanations was correlated with overall ability.

Student certainty might be the key intervening variable here. For example, Miller and Brownwell (1975), in conservation studies involving young children, found that conservers were more willing to press their views than non-

conservers because they had more confidence in the validity of their claims. Tudge (1989, 1990) found that when children were asked to predict the outcome of a balance beam task, those with a decision rule that led to uncertainty in some conditions deferred to children who used a rule that produced certainty in all conditions, regardless of whether the rule was more or less advanced than their own. A critical element in the maintenance of the confidence of students with less powerful but more certain rules was Tudge's decision to give them no feedback on the accuracy of their predictions. Roth and Roychoudhury (1993) observed that senior high school students with greater conviction could persuade their peers, even if they were incorrect, if there was no other information source available. In a similar way, the lack of certainty about correlational reasoning performance described above might have increased the proportion of explanations offered by poorer performers.

Students' ability to give good explanations may have been further hampered by the absence of an appropriate language for describing correlational reasoning concepts. Basili and Sanford (1991) found that small-group discussions in junior college chemistry classes were impeded by students' lack of fluency in using scientific terms. The instructional treatments attempted to provide a few essential technical terms, but the students and teachers did not take to them, substituting fuzzy alternatives (e.g., referring to scale intervals as "places where you break it"). Although previous studies have found that nonverbal behavior is a central element in children's explanations (Bearison, 1982; Cooper, Ayers-Lopez, & Marquis, 1982; Forman, 1989; Mehan & Riel, 1982), it is desirable that it be accompanied by oral commentaries, even in highly visual tasks such as graph construction and interpretation. In giving demonstrations the explanation givers failed to describe their actions in conceptual terms so that the instructions could be generalized from one problem instance to another.

In addition to being flawed in content, the explanations that were offered by students suffered from many of the deficiencies of explanations exhibited by novice teachers. The student explanation givers used a teacher-as-teller strategy in which knowledge is transmitted rather than constructed, an approach characteristic of novice teachers (Florio-Ruane & Lensmire, 1990; Holt-Reynolds, 1992). Leinhardt and Greeno (1986) found that novice teachers were less proficient than experienced practitioners in using questioning to obtain information about student performance that could be used to redirect the lesson. The same difficulty was observed in our study with respect to explanation givers' inability to recover from an unusual help seeker response in a Socratic dialogue. Novice teachers lack the subject knowledge to identify suitable examples (Clermont, Borko, & Krajcik, 1994; Reynolds, 1992). The inability to think of alternate illustrations is the most likely reason why explanation givers in our investigations persisted with examples after the recipients had indicated they were not suitable. Like the students observed by Davis and Sigurdson (1992), our explainers were heavily dependent on the examples provided by teachers and continued with them even when it was clear that the teacher example had produced confusion in the minds of help seekers. There was also evidence that explanation givers reduced the demands of the curriculum by emphasizing completeness over other criteria and by

proposing to help seekers that they adopt simplistic problem definitions. Previous studies have found that these instructional procedures, used by experienced as well as novice teachers, reduce comprehension level tasks to algorithmic or recall assignments and have negative effects on conceptual understanding (Doyle & Carter, 1984; Miller, Leinhardt, & Zigmond, 1988; Sanford, 1985).

Another factor that may have reduced the effectiveness of explanations was help giver insensitivity to the learning needs of their peers. Mehan and Riel (1982) observed that when young children were explaining a game to their peers they usually started with an example provided by the explainer, whereas adult teachers were likely to begin with an example elicited from the child. In our investigations we found the same preference for beginning with the explainer's example, which was exacerbated by an unwillingness or inability to switch to the specific context in which the help seeker raised the problem. This practice placed the cognitive demands of translating the explanation into the desired context on students who may have been overloaded to begin with. Previous investigators (Cooper et al., 1982; Mehan & Riel, 1982) observed that student explainers used more directives than questions and were less likely than adults to obtain feedback on the learner's understanding as the explanation progressed, processes that we observed. Our student explainers employed a teaching strategy that was indifferent to the cognitive structures of learners, perhaps because they were unaware of the need to attend to the perspectives of others. It is also possible that explainer insensitivity to learner needs was a failure of caring. There were several instances of abusive verbal behavior in the recordings. Because previous studies have found that interpersonal behavior in cooperative learning groups changes over time (Ross & Raphael, 1990), it may be that explainer sensitivity might have increased if the experiments had continued over a longer period. Cooper et al. (1982) found that pairs of students who worked together frequently were more likely to share a teaching role.

Explanation Givers Did Not Monitor Help Seekers' Understanding of the Explanations

Hooper (1992) found that the only help giver response that influenced recipients' achievement was checking the helped person's understanding. Peterson and Swing (1985) observed that effective student groups were able to determine when explanations were appropriate and Webb (1989) suggested that student explanations were unlikely to have a beneficial impact unless they were understood and internalized. We observed few instances of help givers explicitly testing whether explanation seekers' needs had been satisfied. Although help givers occasionally inquired whether the recipient understood the explanation, more powerful tests such as asking the help seeker to recapitulate the explanation or apply it to another problem instance were very rare. The infrequency of such probes might be attributed to previously noted problems such as the small number of independent practice tasks assigned, uncertainty of explainers about their own performance, lack of awareness about the need to monitor understanding, and the failure of caring. Students' desire to complete the tasks quickly may also have reduced their willingness to monitor the impact of their explanations: speed of task completion is a criterion of group

success for many cooperative learning students (Holloway, 1990; Mulryan, 1992).

Conclusions

Implications for Teachers

Slavin (1987) argued that the developmental and motivational perspectives on cooperative learning could be reconciled if reward structures stimulated group members to help their peers. Such was not the case in this study. The poverty of explanation might explain why STAD was less effective than Group Investigation on higher-order achievement and more so on lower level objectives (Sharan, Kussell, Bejarano, Raviv, & Sharon, 1985). Two strategies for enhancing STAD come to mind. The first is to train students in interaction processes, as recommended by other cooperative learning approaches. The second is to select better examples for illustrating new ideas that are introduced in the teacher-directed phase of the lesson.

The training strategy might consist of the teacher giving explicit models of explanation in which the processes of asking for an explanation, giving one, and monitoring its impact could be modeled. Swing and Peterson (1982) found that the frequency and quality of explanations offered by primary age children could be enhanced through role playing that provides exemplars of good and poor explanations. Webb, Qi, Yan, Bushey, and Farivar (1990) found that role playing when combined with related strategies contributed to grade 7 students' ability to explain how to solve math problems. In another study we found that giving students feedback on their conversations had a positive impact. When students were given edited transcripts of their talk, along with teacher models of productive explanation, the frequency and quality of student interactions improved (Ross, in press). We also found that the teacher's interventions when students were working in groups influenced the explanations they gave to one another (Ross, 1994). The frequency and quality of students' explanations to their peers can also be enhanced through training in reciprocal questioning techniques in which students alternate between listener and critic (King, 1990; O'Donnell & Dansereau, 1992).

A different, but not mutually exclusive, strategy might consist of teachers selecting as concrete illustrations rich cases that optimize links between new ideas and existing understanding. The teachers in Davis and Sigurdson's (1992) study selected the examples used in whole-class demonstrations on the basis of their own out-of-school interests rather than on compatibility with student interests and experiences. Of particular concern are analogies that lend themselves to the development of misconceptions (Clement, Brown, & Zietsman, 1989, referred to these as "brittle anchors") because student explanations tend to be based on analogies to everyday experience (Hesse & Anderson, 1992). Because teacher illustrations are likely to be repeated by student explainers, contexts that create confusion, such as misleading analogies, are likely to persist in student discussions.

Finally, the failure of students to seek help when they need it might be reduced if teachers provided students with unambiguous appraisal criteria. Such criteria would enable those who need help to recognize it and to be able to identify peers with the academic ability to provide effective explanations.

Implications for Researchers

The findings from this research suggest several fruitful lines of investigation. The most interesting is to examine the impact of treatment duration on the effectiveness of explanation giving and receiving. Duration is likely to have an impact by increasing students' commitment to one another (although it should not be assumed that increased interaction will necessarily improve group cohesion) and by increasing students' familiarity with the skills and learning needs of their peers. Cooperative interactions over extended time periods might also result in explanation givers recognizing differences among students, becoming aware, for example, that not all students learn in the same way as oneself. Kagan (1992) concluded from a review of research on teacher growth that through practice teaching novice teachers reconstruct their images of teaching, and ultimately their practice, by obtaining knowledge of the variability of students. Increased duration might further impact explanation seeking by contributing to the development of a shared language, both verbal (Buckholdt & Wodarski, 1978) and nonverbal (Allan & Feldman, 1973).

Notes

1. Funding for this research was provided by the Ontario Ministry of Education through its Block Transfer Grant to the Ontario Institute for Studies in Education (#81-1118) and the Social Sciences and Humanities Research Council (#410-90-0834). The views expressed in the article do not necessarily reflect the views of the Ministry or SSHRC. Deborah Berrill contributed to the data collection in sample (a) and Anne Hogaboam-Gray contributed to the data collection in samples (b) and (c).
2. The software program was designed to support the correlational reasoning instruction. In addition to file management procedures, it contained simple commands to create multivariate scatterplots and calculate correlation coefficients (Hogaboam-Gray, Ross & Cousins, 1991).

References

Allan, V., & Feldman, R. (1973). Learning through tutoring: Low achieving children as tutors. *Journal of Experimental Education, 42*, 1-5.

Azmitia, M. (1988). Peer interaction and problem solving: When are two heads better than one? *Child Development, 59*(1), 87-96.

Bargh, J.A., & Schul, Y. (1980). On the cognitive benefits of teaching. *Journal of Educational Psychology, 72*, 593-604.

Basili, P., & Sanford, J. (1991). Conceptual change strategies and cooperative group work in chemistry. *Journal of Research in Science Teaching, 28*, 293-304.

Bearison, D. (1982). New directions in studies of social interaction and cognitive growth. In F. Serafica (Ed.), *Social cognitive development in context* (pp. 199-221). New York: Guilford.

Brown, J., Collins, A., & Diguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*, 32-42.

Buckholdt, D., & Wodarski, J. (1978). The effects of different reinforcement systems on cooperative behaviors exhibited by children in classroom contexts. *Journal of Research and Development in Education, 12*, 50-68.

Clement, J., Brown, D., & Zietsman, A. (1989). Not all preconceptions are misconceptions: Finding "anchoring conceptions" for grounding instruction on students' intuitions. *International Journal of Science Education, 11*, 554-565.

Clermont, C., Borko, H., & Krajcik, J. (1994). Comparative study of the pedagogical content knowledge of experienced and novice chemical demonstrators. *Journal of Research in Science Teaching, 4*, 149-441.

Cooper, C., Ayers-Lopez, S., & Marquis, A. (1982). Children's discourse during peer learning in experimental and naturalistic situations. *Discourse Processes, 5*, 177-191.

Corno, L. (1989). Self-regulated learning: A volitional analysis. In B. Zimmerman & D. Schunk (Eds.), *Self-regulated learning and academic achievement* (pp. 111-142). New York: Springer-Verlag.

Corno, L. (1992). Encouraging students to take responsibility for learning and performance. *Elementary School Journal, 93*(1), 69-83.

Cousins, J.B., & Ross, J. (1993). Improving higher order thinking skills "with" the computer: A comparative study. *Journal of Research on Computing in Education, 26*(1), 94-115.

Davis, B., & Sigurdson, S. (1992). *Talk in the mathematics classroom: Implementation and research concerns*. Paper presented at CSSE, Charlottetown.

Deering, P., & Meloth, M. (1990). *An analysis of the content and form of students' verbal interactions in cooperative groups*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Doise, W., & Mugny, G. (1979). Individual and collective conflicts of centrations in cognitive development. *European Journal of Social Psychology, 9*, 105-108.

Doyle, W., & Carter, K. (1984). Academic tasks in classrooms. *Curriculum Inquiry, 14*, 129-149.

Florio-Ruane, S., & Lensmire, T. (1990). Transforming future teachers' ideas about writing instruction. *Journal of Curriculum Studies, 22*, 277-289.

Forman, E. (1989). The role of peer interaction in the social construction of mathematical knowledge. *International Journal of Educational Research, 13*, 55-70.

Good, T., Slavings, R., Harel, K., & Emerson, H. (1987). Student passivity: A study of question asking in K-12 classrooms. *Sociology of Education, 60*, 181-199.

Hertz-Lazarowitz, R. (1989). Cooperation and helping in the classroom: A contextual approach. *International Journal of Educational Research, 13*, 113-119.

Hesse, J., & Anderson, C. (1992). Students' conceptions of chemical change. *Journal of Research in Science Teaching, 29*, 277-299.

Hogaboam-Gray, A., Ross, J.A., & Cousins, J.B. (1991). *CORReoGRAPH—A program for solving correlational reasoning problems: User's guide*. Peterborough, ON: OISE Trent Valley Centre.

Holloway, S. (1990). *Contradictory meanings in teaching/learning "cooperation" via cooperative learning*. Paper presented at the annual meeting of the American Educational Research Association, Boston.

Holt-Reynolds, D. (1992). Personal history-based beliefs as relevant prior knowledge in course work. *American Educational Research Journal, 29*, 325-349.

Hooper, S. (1992). Effects of peer interaction during computer-based mathematics instruction. *Journal of Educational Research, 85*, 180-189.

Johnson, D. (1990). *Reaching out: Interpersonal effectiveness and self-actualization* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Johnson, D., & Johnson, R. (1987). *Learning together and alone: Cooperative, competitive, and individualistic learning*. Englewood Cliffs, NJ: Prentice-Hall.

Johnson, R., Johnson, D., & Stanne, M. (1986). Comparison of computer-assisted cooperative, competitive, and individualistic learning. *American Educational Research Journal, 23*, 382-392.

Kagan, D. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research, 62*, 129-169.

Kagan, S. (1988). *Cooperative learning: Resources for teachers*. Riverside, CA: University of California.

Karenbenick, S., & Knapp, J. (1988). Help seeking and the need for academic assistance. *Journal of Educational Psychology, 80*, 406-408.

Kempa, R., & Ayob, A. (1991). Learning interactions in group work in science. *International Journal of Science Education, 13*, 341-354.

King, A. (1989). Verbal interaction and problem-solving within computer-assisted cooperative learning groups. *Journal of Educational Computing Research, 5*(1), 1-15.

King, A. (1990). Enhancing peer interaction and learning in the classroom through reciprocal questioning. *American Educational Research Journal, 27*, 664-687.

Lawson, A., & Bealer, J. (1984). The acquisition of basic quantitative reasoning skills during adolescence: Learning or development. *Journal of Research in Science Teaching, 21*, 417-423.

Leinhardt, L., & Greeno, J. (1986). The cognitive skill of teaching. *Journal of Educational Psychology, 78*, 75-95.

Markman, E. (1977). Realizing that you don't understand: A preliminary investigation. *Child Development, 48*, 986-992.

Markman, E. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development, 50*, 643-655.

Mehan, H., & Riel, M. M. (1982). Teachers' and students' instructional strategies. In L. Adler (Ed.), *Cross cultural research at issue* (pp. 171-188). New York: Academic Press.

Miller, S., & Brownwell, C. (1975). Peers, persuasion and Piaget: Dyadic interaction between conservers and nonconservers. *Child Development, 46*, 992-997.

Miller, S., Leinhardt, G., & Zigmond, N. (1988). Influencing engagement through accommodation: An ethnographic study of at-risk students. *American Educational Research Journal, 25*, 465-487.

Mulryan, C. (1992). Student passivity during cooperative small groups in mathematics. *Journal of Educational Research, 85*, 261-273.

Newman, R. (1990). Children's help-seeking in the classroom: The role of motivational factors and attitudes. *Journal of Educational Psychology, 82*, 71-80.

Newman, R., & Goldin, L. (1990). Children's reluctance to seek help with schoolwork. *Journal of Educational Psychology, 82*, 92-100.

Newman, R., & Schwager, M. (1993). Students' perceptions of the teacher and classmates in relation to reported help seeking in math class. *Elementary School Journal, 94*, 3-17.

O'Donnell, A., & Dansereau, D. (1992). Scripted cooperation in student dyads: A method for analyzing and enhancing academic learning and performance. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 120-141) Cambridge, UK: Cambridge University Press.

Ontario Ministry of Education. (1988). *Intermediate and senior geography curriculum guidelines*. Toronto, ON: Queen's Park.

Ontario Ministry of Education. (1992). *Policy framework for the transition years*. Unpublished manuscript. Toronto, ON: Queen's Park.

Palincsar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-monitoring activities. *Cognition and Instruction, 5*, 269-287.

Peterson, P.L., & Swing, S.R. (1985). Students' cognitions as mediators of the effectiveness of small-group learning. *Journal of Educational Psychology, 77*, 299-312.

Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66*, 211-227.

Reynolds, A. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research, 62*, 1-36.

Ross, J.A. (1994). *Explaining solution strategies: a case study of student talk in a math reform classroom.* Paper presented at the annual meeting of the Canadian Society for the Study of Education, Calgary.

Ross, J.A. (in press). Effects of feedback on student behavior in cooperative learning groups in a grade 7 math class. *Elementary School Journal.*

Ross, J.A., & Cousins, J.B. (1993a). Patterns of student growth in reasoning about multivariate correlational problems. *Journal of Educational Psychology, 85*, 1-17.

Ross, J.A., & Cousins, J.B. (1993b). Enhancing secondary school students' acquisition of correlational reasoning skill. *Research in Science and Technological Education, 11*(3), 191-206.

Ross, J.A., & Cousins, J.B. (in press). The impact of explanation seeking on student achievement and attitudes. *Journal of Educational Research.*

Ross, J.A., & Raphael, D. (1990). Communication and problem solving achievement in cooperative learning groups. *Journal of Curriculum Studies, 22*, 149-164.

Roth, W., & Roychoudhury, A. (1993). The concept map as a tool for the collaborative construction of knowledge: A micro-analysis of high school physics students. *Journal of Research in Science Teaching, 30*, 503-534.

Sanford, J. (1985). *Comprehension-level tasks in secondary classrooms. R&D Report No. 6199.* Austin, TX: R&D Centre for Teacher Education.

Sharan, S., Kussell, P.H.R., Bejarano, Y., Raviv, S., & Sharan, Y. (1985). Cooperative learning effects on ethnic relations and achievement in Israeli junior-high school classrooms. In R. Slavin, S. Sharan, S. Kagan, R. Hertz-Lazarowitz, C. Webb, & R. Schmuck (Eds.), *Learning to cooperate, cooperating to learn* (pp. 313-344). New York: Plenum Press.

Sharan, S., & Shachar, H. (1988). *Language and learning in the cooperative classroom.* New York: Springer Verlag.

Slavin, R. (1983). When does cooperative learning increase student achievement? *Psychological Bulletin, 94*, 429-445.

Slavin, R. (1987). Developmental and motivational perspectives on cooperative learning: A reconciliation. *Child Development, 58*, 1161-1167.

Slavin, R. (1990). *Cooperative learning: Theory, research, and practice.* Englewood Cliffs, NJ: Prentice-Hall.

Slavin, R. (1992). When and why does cooperative learning increase achievement: Theoretical and empirical perspectives. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 145-173). Cambridge, UK: Cambridge University Press.

Solomon, J. (1992). The classroom discussion of science-based social issues presented on television: Knowledge, attitudes and values. *International Journal of Science Education, 14,* 431-444.

Swing, S., & Peterson, P. (1982). The relationship of student ability and small-group interaction to student achievement. *American Educational Research Journal, 19,* 259-274.

Tudge, J. (1989). When collaboration leads to regression: Some negative consequences of socio-cognitive conflict. *European Journal of Social Psychology, 19,* 123-189.

Tudge, J. (1990). *Collaborative problem solving in the zone of proximal development.* Paper presented at the annual meeting of the American Educational Research Association, Boston.

Webb, N. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research, 13,* 21-39.

Webb, N. (1992). Testing a theoretical model of student interaction and learning in small groups. In R. Hertz-Lazarowitz & N. Miller (Eds.), *Interaction in cooperative groups: The theoretical anatomy of group learning* (pp. 102-119), Cambridge, UK: Cambridge University Press.

Webb, N., & Kenderski, C. (1984). Student interaction and learning in small group and whole class settings. In P. Peterson, L. Wilkinson, & M. Hallinan (Eds.), *The social context of instruction: Group organization and group processes* (pp. 153-170). New York: Academic Press.

Webb, N., & Kenderski, C. (1985). Gender differences in small-group interaction and achievement in high- and low-achieving classes. In L. Wilkinson & C. Marrett (Eds.), *Gender influences in classroom interaction* (pp. 209-236). Orlando, FL: Academic Press.

Webb, N., Qi, S., Yan, K., Bushey, B., & Farivar, S. (1990). *Cooperative small-group problem solving in middle school mathematics.* Paper presented at the American Educational Research Association, Boston.

Wilkinson, L., & Calculator, S. (1982). Requests and responses in peer-directed reading groups. *American Educational Research Journal, 19,* 107-120.

Wilkinson, L.C., & Spinelli, F. (1983). Using requests effectively in peer-directed instructional groups. *American Educational Research Journal, 20,* 479-501.

Yap, K., & Yeany, R. (1988). Validation of hierarchical relationships among Piagetian cognitive modes and integrated science process skills for different cognitive reasoning levels. *Journal of Research in Science Teaching, 25,* 247-281.

Book Reviews

Genie: An Abused Child's Flight from Silence. Russ Rymer. New York: Harper Collins, 1993. 228 pages. ISBN: 0-06-016910-9.

Reviewed by *Tracey Derwing, University of Alberta*

This is the first book with a linguistics issue as a central theme that has ever moved me to tears. I was already familiar with much of the content, through an in-person account from one of the peripheral players involved, through Susan Curtiss's (1977) book on Genie's language development, and through Rymer's extensive article in the *New Yorker* on which this book is based. Even so, it is impossible to develop an immunity to the awful miscommunications on the part of people who had as their goal the rehabilitation of a child and the ramifications of the treatment she received. The story of Genie is a cautionary tale for researchers and educators; if ever there was a situation that exemplified the expression "good intentions pave the way to hell," this is it.

Genie was 13 years old when her mother took her and escaped the family home, where Genie's father had kept Genie captive, tied to a potty chair in a barren room for most of her life. The child was physically and cognitively underdeveloped; she had been spoken to only rarely and had not developed more than a few words (her productive vocabulary consisted of *stopit* and *nomore*). Rymer's book is a chronicle of what happened after Genie's plight was discovered in November of 1970 to the present. Along the way he discusses the nature of language acquisition, in particular the ongoing debate between those who believe that language is innate and those who maintain that it must be learned through interaction. Genie was thought to be an important piece of the puzzle: according to Chomsky's linguistic theory, the innate language acquisition device (latterly known as Universal Grammar) deteriorates over time, thus humans should not be able to learn a first language after the onset of puberty. If Genie succeeded in learning English, she would disprove the "critical period" hypothesis; she would also provide empiricists with strong evidence for the need for interaction in order to develop language.

The "Genie team," as Rymer dubbed them, was the group of doctors, psychologists, psychiatrists, linguists, and teachers who were to work with the girl. Rymer describes the first major meeting in which the scientists determined what the focus of their research with Genie should be. Interestingly, the evening before the conference, the Genie team and other experts viewed François Truffaut's film *The Wild Child* based on the diaries of physician Jean-Marc-Gaspard Itard who worked with Victor, another child who grew up in virtual isolation. Although Victor was discovered in 1800, the parallels with Genie are remarkable. The scientists were struck by the film; they shared the optimism that Itard's notes expressed with regard to Victor's progress. As Rymer indicates, however, Truffaut filmed only the first years of Itard's journals; Rymer deftly outlines the similarities between Genie and Victor later on in these two

natural experiments—Victor's slide from celebrity to obscurity, from constant testing and attention to benign neglect. Although there were experts at the Genie conference who expressed some concerns about the direction the research was to take, namely, Genie's linguistic development, those most closely involved were convinced that what was in the best interests of research was also in the best interests of the child.

Rymer extensively interviewed as many of the principals as possible; he also obtained court records, personal letters, and other documentation. He immersed himself in linguistic theory (both sides of the fence) in order to present the whole story as fairly as he could. The only person with whom he had no contact (either direct or indirect) is Genie herself, yet she is the heart and soul of the book. Nearly all undergraduate linguistics students know that Genie was a most compelling communicator; they have heard anecdotes about passersby on the street handing her a dollie or a necklace or some other object—all reporting that the little girl had somehow let them know that she wanted it. Susan Curtiss, who spent years with Genie, initially tracing her language development but eventually becoming her friend and supporter, is quoted as saying "Genie is the most powerful, most inspiring person I have ever met. I'd give up my job, I'd change careers to see her again" (p. 221). To Rymer's credit, Genie has communicated here as well.

Rymer outlines the events surrounding Genie's initial stay at Children's Hospital of Los Angeles, her placement in the home of her teacher, her four-year stay in the home of her psychologist, her short reunification with her mother, and a series of foster homes after that. There were misunderstandings and disagreements from the start as to what Genie needed and who would best provide for her needs. Disagreement turned to acrimony and vendetta; a confounding of roles of researcher and foster parent led to disaster. The two individuals who are most perplexing in the sordid details of the book are Jean Butler Ruch and David Rigler. Jean Butler was Genie's teacher; she clearly loved Genie and wanted to keep her in her home. Although she was officially a part of the Genie team, evidently Butler was not a team player. She had had a record of difficulty working with others, and she evidenced similar problems in her role at Children's Hospital. When she was denied foster parent status by the Department of Public Social Services, she started a campaign against those who were associated with the Genie team, writing letters to other scientists, castigating the team's efforts. She also befriended Genie's biological mother and eventually convinced her to sue nearly everyone involved in the case. Rigler, Genie's psychologist, seems to have been largely motivated by academic ambition and greed, yet he and his family sacrificed a great deal by having Genie in their home for four years, where she was apparently making good progress. Within a month of hearing of a failure to receive renewal on an NIMH grant, however, Rigler sent Genie back to her biological mother.

In an interview with Jay Shurley, a psychiatrist who was a consultant from the outset on Genie's case and one of the only people who is still allowed to see Genie, it is revealed that she is now in an institution. "The way I think of Genie, she was this isolated person, incarcerated for all those years, and then she emerged and lived in a more reasonable world for a while, and responded to

this world, and then the door was shut and she withdrew again and her soul was sick" (p. 213).

The shutting of the door is the crucial point of this book. Although all involved were convinced of the ethical nature of their research, it is the bungling of ethical issues, the failure to see Genie as a whole, a person who will no doubt live for some years to come, that resulted in her regression. The progress that was so carefully documented in test after test after test, and years of intense observation has been lost. There was no long-term plan for Genie, no integrated approach to her rehabilitation. The expectations of her were too high—she was to solve a question that has not been resolved for centuries; when she didn't provide an answer, she was shunted off to an institution through a litigious and bureaucratic route.

What Rymer has provided in this book is a complete picture of the complicated mess that was made of Genie's treatment and the very limited knowledge gleaned about language acquisition as a result of her study. Although some linguists claim that Genie's failure to acquire grammar supports the innatist stand, interactionists such as Catherine Snow ask, "How could a child who lacked language because she had been shut away from her mother be proof of the contention that our mothers don't teach us language?" (p. 157).

The book has a few oversimplifications with regard to linguistic matters that are mildly irritating, but none that affect the basic thesis. It is unfortunate that no references are listed. This is intended to be a book for the layperson; however, the author owes the reader original source information. These complaints pale in light of the overall impression. The highly readable style, the careful documentation of detail, and the fascinating nature of the subject matter make this a thought-provoking and thoroughly engaging account of a tragedy.

Reference

Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day "wild child."* New York: Academic Press.

Tracey Derwing is an associate professor who teaches teaching English as a second language. Her research interests include native speaker-nonnative speaker discourse, adult second language literacy, and citizenship education for immigrants.

Violence and Abuse in the Lives of People With Disabilities. Dick Sobsey. Baltimore: Paul H. Brookes, 1994.

Reviewed by *Gregor Wolbring, German Council of Self-Determined Living Centres*

I am pleased to review the book *Violence and Abuse in the Lives of People With Disabilities* by Dick Sobsey. In order to place the tone of my review in context, I should state that I am not from the educational field by profession; I am a biochemist. Nonetheless, as a disability activist, I feel confident to review this book.

Although several books are available on the topic of abuse and disabilities (e.g., Garbarino, Brookhouser, & Authier, 1987; Morgan, 1987; Senn, 1987), Sobsey's 1994 book is the first to provide a comprehensive model that helps us understand abuse and other forms of maltreatment (e.g., euthanasia has previously been seen as an independent phenomenon).

As a disability activist, I feel that more should be written about the violence against people with disabilities. Drawing on previous work (Sobsey, Gray, Wells, Pyper, & Reimer-Heck, 1991; Sullivan, Brookhauser, Scanlan, Knutson, & Schulte, 1991; Turk & Brown, 1992; Westcott, 1993), Sobsey outlines in a clear and detailed fashion the probability that a disabled person will be abused by a family member (16.5%), an acquaintance or neighbor (16.5%), a disability service provider (28%), or a stranger (6.6%). In addition the book clearly describes how the risk increases that people with a mental disabilities will be victimized. Using data from Wilson and Brewer (1992), Sobsey suggests that mentally disabled people are more than 10 times as likely as other citizens to be victims of robbery or sexual assault.

Violence and Abuse in the Lives of People With Disabilities shows us how society trivializes or decriminalizes offenses against disabled people. For example, rather than describe an act of murder of a disabled person as murder, we use terms such as euthanasia or assisted suicide (for further discussion of this issue, see Luckasson, 1992).

The tables and statistics in the book are powerful reminders of the frequency and severity of violence against people with disabilities that should make readers shudder with disgust. *Violence and Abuse in the Lives of People With Disabilities* will surely generate greater demand for laws and standards that protect and integrate disabled people. Equally powerful is the discussion of Nazi history and the presentation of an ecological model of abuse.

Perhaps the most compelling part of the book is the elaborate section (nearly half the book) that describes how abuse can be prevented. This section is divided into several topics: Empowering individuals to resist abuse; Families and other caregivers: support and selection; Building safer environments; Law and law enforcement; Changing attitudes that disinhibit violence; Healing the consequences of abuse; and Prevention and intervention teams. I would like to single out the topic "Changing attitudes that disinhibit violence" for discussion in this review. This section clearly shows the "Catch 22" situation that people with disabilities face. Society, especially through the media, presents pictures of disabled people that can best be understood as two stereotypes. Either disabled people are portrayed as suffering individuals who are a burden to society or they are seen as superheroes. The stereotype of the suffering individual who is a burden to society is used a lot these days in an attempt to push for genetic testing, assisted suicide, or euthanasia ("mercy killing" is the phrase often used here). The stereotype of superhero portrays the disabled person as productive against all odds. Both stereotypes lead to a situation where people with a disability are viewed and treated differently than others. Being treated differently leads disabled people to behave differently; it also threatens their self-esteem. Their own different behavior often acts to confirm the stereotypes that people hold. Having said this, I feel that the part of

Sobsey's book that deals with attitudes toward people with a disability and how attitudes might be changed is of utmost importance.

To summarize, *Violence in the Lives of People With Disabilities* extends previous work with an integrated ecological model of violence that includes elements from social labeling theory and ambivalence theory. From this perspective, child abuse and sexual assault can no longer be seen as unique phenomena, but rather as part of the larger spectrum of negative outcomes that result from discriminatory attitudes.

References

Garbarino, J., Brookhouser, P. E., & Authier, K.J. (Eds.). (1987). *Special children-special risks: The maltreatment of children with disabilities*. New York: Aldine de Gruyter.

Luckasson, R. (1992). People with mental retardation as victims of crime. In R.W. Conley, R. Luckasson, & G.N. Bouthilet (Eds.), *The criminal justice system and mental retardation: Defendants and victims* (pp. 209-220). Baltimore, MD: Brookes.

Morgan, S.R. (1987). *Abuse and neglect of handicapped children*. Boston, MA: College Hill Press.

Senn, C.Y. (1987). *Vulnerable: Sexual abuse and people with an intellectual handicap*. Downsview, ON: G. Allan Roeher Institute. (ERIC Document Reproduction Service No. ED 302 975)

Sobsey, D., Gray, S., Wells, D., Pyper, D., & Reimer-Heck, B. (1991). *Disability, sexuality, and abuse: An annotated bibliography*. Baltimore, MD: Brookes.

Sullivan, P.M., Brookhauser, P.E., Scanlan, J.M., Knutson, J.F., & Schulte, L.E. (1991). Pattern of physical and sexual abuse of communicatively handicapped children. *Annals of Otology, Rhinology, and Laryngology, 100*(3), 188-194.

Turk, V., & Brown, H. (1992, August 5-7). *Sexual abuse of adults with learning disabilities*. Paper presented at the Conference of the International Association for Scientific Study of Mental Deficiency, Brisbane, Australia.

Westcott, H. (1993). *Abuse of children and adults with disabilities*. London: National Society for Prevention of Cruelty to Children.

Wilson, C., & Brewer, N. (1992). The incidence of criminal victimization of individuals with an intellectual disability. *Australian Psychologist, 2*, 114-117.

Gregor Wolbring is a research scientist in the biochemistry department at the University of Calgary. As a disability activist he has delivered several invited presentations on disability-related topics, genetics, reproductive technologies, ethics, assisted suicide, euthanasia, and the interaction between the women's movement and the disability movement.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the
References under the heading *Notes;* any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
845 Education Centre South
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1438



THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XLI NUMBER 2 JUNE 1995

The Alberta Journal of Educational Research

Published in March, June, September, and December by the Faculty of Education, University of Alberta.

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
Mount Allison University

George Buck
University of Alberta

Ardra L. Cole
Ontario Institute for Studies in Education

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Heather Ryan
Massey University

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Ken Ward
Robert H. Short

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, the Alberta Advisory Committee for Educational Studies, and the Research Grants Office of the University of Alberta

The subscription rate is $32.00 per year for individuals, $45.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Office of the Dean
845 Education Centre South
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0236

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XLI, Number 2, June 1995

Articles

W. Rod Dolmage **127** Accusations of Teacher Sexual Abuse of Students in Ontario Schools: Some Preliminary Findings

K. Vern Stenlund **145** Teacher Perceptions Across Cultures: The Impact of Students on Teacher Enthusiasm and Discouragement in a Cross-cultural Context

Trevor J. Gambell **162** Ethnography as Veneration

Michael W. Pratt
Tania Filipovich
Marie Bountrogianni **175** Teachers' Views of Parents: Family Decision Making Styles and Teacher-Parent Agreement Regarding Homework Practices and Values

David R. Jones
Edward A. Holdaway **188** Expectations Held for Department Heads in Postsecondary Institutions

Myrna L. Greene
Carillon R. Purvis **213** The Forgotten Link: Transitions from Graduate School to Classroom Teaching

Book Review

Diane G. Symbaluk **231** Women's Work is Never Done: Resolving Gender Inequalities in Education

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 127-144*

W. Rod Dolmage
University of Western Ontario

Accusations of Teacher Sexual Abuse of Students in Ontario Schools: Some Preliminary Findings

Issues related to accusations of teacher sexual abuse of students seem to be among those subjects about which there is no lack of opinionated discussion, even though few of the "facts" on which the opinions are based are rooted in verifiable research. Reports in the media would appear to describe an epidemic of teacher sexual exploitation and assault of students in the public schools. On the other hand, teachers' federations claim that although accusations of sexual abuse may be on the increase, legitimate cases of teachers sexually exploiting or abusing students are relatively rare. The article that follows reports some preliminary results of attempts to establish baseline data concerning the prevalence of accusations, charges, and convictions, and suggests that much more data are needed before definitive conclusions should be drawn or meaningful recommendations can be made.

Les questions concernant les accusations d'abus sexuels commis par les enseignants et enseignantes dont les victimes sont les élèves semblent susciter de nombreuses discussions intransigeantes fondées sur très peu de faits provenant rarement de recherches vérifiables. Les reportages des médias semblent décrire une épidémie d'exploitation et d'agressions sexuelles commises par les enseignants et enseignantes envers les élèves dans des écoles publiques. Dautre part, les fédérations et les syndicats professionnels d'enseignants prétendent que les cas légitimes d'exploitations et d'abus sexuelles de la part des enseignants et enseignants envers les élèves sont relativement rares même si ces accusations semblent croître en nombre et en fréquences. L'article qui suit présente des résultats préliminaires de tentatives d'établir une ligne de base de données des fréquences de plaintes d'accusations, de condamnations d'abus, et d'exploitations sexuelles. L'article suggère qu'un plus grand nombre de données soient rassemblées afin de pouvoir en tirer des conclusions définitives et de proposer des recommandations significatives.

The Problem

TEACHER ACCUSED OF MOLESTING 11 GIRLS IN CLASS. (1991)
LESBIAN TEACHER ACCUSED OF SEX WITH STUDENTS. (1992)
CLEARED IN SEX CASE, TEACHER MAY QUIT JOB. (1987)
TORONTO TEACHER CLEARED IN SEXUAL ABUSE CASE. (1985)

These are but a few of many similar headlines that seem all too frequently to confront Canadian newspaper readers. To the average citizen headlines like these might well appear to describe an epidemic of teacher sexual exploitation and assault of the students entrusted to their care. On the other hand, teachers' federations claim that although accusations of sexual abuse may appear to be

Rod Dolmage is an associate professor in the Division of Educational Policy Studies, Faculty of Education. His research and teaching interests include teacher welfare, educational law, and the politics of education.

on the increase, legitimate cases of teachers sexually exploiting or abusing students are relatively rare.

Discrepancies of this magnitude raise some obvious questions. It would clearly be useful to know, for example, how many teachers are accused of sexually abusing students in a given jurisdiction in a given year, how many accused teachers are actually charged under the Criminal Code of Canada (1985) and how many are convicted. Given this information, one might be in a position, taking into account the size of the educational enterprise in Canada, to consider whether the numbers actually represent an "epidemic" of sexual abuse in schools. In addition, given that the consequences of an accusation of sexual abuse (whether substantiated or not) can be devastating for the teacher and for the teaching profession, it would be useful to know how many teachers are falsely accused of sexually abusing their students.[1]

One might have thought the answers to such questions would be fairly easy to discover. Surprisingly, at least to me, answers to these basic questions are not readily available. Indeed it would appear that issues related to accusations of teacher sexual abuse of students are among those subjects about which there is no lack of opinionated discussion, regardless of the fact that almost no one seems to have any basic data concerning the phenomenon or its implications. To further complicate matters, those who do have data are reluctant to share them.

One would expect that the academic literature would shed some light on these issues. The psychological and legal literature offers extensive debate on the question of the existence or nonexistence of a "false" or "repressed" memory syndrome related to child sexual abuse occurring in the distant past (Byrd, 1994; Gleaves, 1994; Gold, Hughes, & Hohnecker, 1994; Loftus, 1993, 1994; Olio, 1994; Peterson, 1994; Rabinovici, 1993; Shorten, 1994). There are some obvious similarities between the false long-term memories allegedly implanted by overzealous therapists such as those discussed by Loftus (1994) and the false short-term memories allegedly created in the minds of nursery school children by overzealous police officers and children's aid workers (Rabinowitz, 1993). However, few such reports relate to sexual abuse in public schools; thus the debate does little to address the issues raised in this article.

Anecdotal accounts such as those illustrated by the headlines cited at the beginning of this discussion appear with regularity in the popular press. American "educational trade magazines" (as opposed to academic journals) have run various articles attempting to define the problem (DeMitchell, 1981), advocating methods of screening teachers and potential teachers (Natale, 1993; Ross, 1985; Zakariya, 1988), warning school boards of their liability if they fail to protect students and employees from sexual harassment and abuse (Marczely, 1993), and describing the aftermath of false accusations (Wilson & Littleson, 1987).

Canadian teachers' federation publications have featured articles on the issues associated with the phenomenon ("Abuse: An Agonizing Issue," 1990; Cooney, Gardhouse, & Getty, 1988; Kendall, 1990; O'Connor, 1989; Stewart, 1989a, 1989b; Sundby, 1990; Ulrich, 1989). Most of these articles focus on how teachers can avoid situations in which they might be accused. Canadian teachers have been warned by their federations and others to exercise extreme care

when touching students, regardless of apparent justification (Polanyi, 1988; Stewart, 1989b; Yaworsky, 1992). Recently, Doctor (1994) described four types of sexual misconduct cases that occur in schools and identified two types of sexual offenders. She also recommends that Canadian faculties of education, provincial certification agencies, and school boards institute criminal background checks of applicants.

Almost all of the "data" cited in these articles are anecdotal or derived from anecdotal evidence, and thus do little to inform us concerning the actual incidence or accuracy of sexual abuse accusations against Canadian public schoolteachers.

Closer to the point, the *Report of the Committee on Sexual Offenses Against Children and Youths* (1984) indicated that one in two women and one in three men would suffer at least one unwanted sexual act in her or his life. Further, the report suggested that four out of five of these incidents would happen before the individual reached the age of 21 (Stewart, 1989a). However, the *Report* (1984, p. 217) also pointed out that studies of the relationship between child sexual assault victims and suspected assailants indicated that only 1 to 5.3% of assailants came from the "position of trust" category that included teachers as well as day care workers, doctors, social workers, school bus drivers, school crossing guards, Big Brother/Big Sister youth workers, ministers/priests/rabbis, camp counselors, dentists, and so forth. Teachers accounted for 10% of alleged assailants from the "position of trust" category (1984, p. 526). The Ontario Ministry of Community and Social Services reported that in the five years between 1983 and 1988, registered child sexual abuse cases increased by 360% from 601 to 2,152 cases per year; however, charges were laid in less than half of these cases (Flavelle, 1988). Neither the number of eventual convictions nor the number of teachers allegedly involved was indicated. More recently, Trocmé, McPhee, Tam and Hay (1994) published the results of their study of reported child abuse and neglect in Ontario. They found that teachers were the alleged perpetrators of sexual abuse in 45 (.4%) of the 11,307 sexual abuse investigations they studied. None of these 45 accusations was substantiated following investigation (p. 69).[2]

"Teachers to Sue" (1988) stated that "the number of child-abuse charges against Ontario teachers increased to more than 80 during the past school year [1987-1988], from 18 in 1983." We have no way of knowing how many of these charges were eventually substantiated. Another newspaper article published earlier the same year (Teahen, 1988) stated that false accusations of student sexual abuse by teachers increased 20% over the same five-year period. With the exception of the *Report* (1984) and Trocmé et al. (1994), none of these data can be attributed to academic sources. Nevertheless, conclusions are being drawn and recommendations are being made. Obviously research that can provide insight into the answers to basic questions such as those outlined above is badly needed.

Method

It was with such basic questions in mind that a research design was developed to gather data concerning phenomena related to student accusations of teacher sexual abuse. Two strategies were used to gather data. The first strategy involved polling Ontario Teachers' Federation (OTF) Affiliates[3] concerning the

experience of their membership over a five-year period. The second strategy employed a questionnaire to gather data from a small sample of experienced Ontario administrators concerning their experience with the phenomenon over a five-year period. Both phases of the investigation and the data obtained are described in the following sections.

Study 1: Data Provided by OTF Affiliates

The first research strategy involved polling Ontario Teachers' Federation Affiliates. In 1991-1992 each of the five Affiliates was contacted by mail and asked to participate in a study that would address a number of questions related to accusations of teacher sexual abuse of students; specifically, each was asked (a) to provide data concerning actual numbers of teachers reporting accusations of sexual abuse in each of the most recent five school years, and (b) to provide information concerning the actual outcomes of the investigations that followed from these accusations (i.e., numbers of teachers charged, numbers convicted or acquitted, numbers of convictions or acquittals appealed, results of appeals, etc.). This design, although simple enough, encountered two significant obstacles.

First, federation officers pointed out that many, perhaps most, cases in which students accuse teachers of sexual abuse are dealt with at the local level. It would appear that unless investigation of the accusation suggests that further steps are necessary (e.g., police involvement or school board action), many such accusations are handled by the school's administration and local federation affiliate representatives without any formal federation involvement. Because the affiliates' central offices may or may not be officially notified concerning many of these incidents, they have no accurate records that can reveal how many such accusations occur. However, if a criminal charge is laid, the relevant affiliate is almost always asked to provide assistance to the accused teacher. Obviously the affiliates maintain records of their activities in this regard. Therefore, although it would not be possible to compile data concerning numbers of teachers accused of sexually abusing students, it would be possible to determine fairly accurately how many teachers were charged with sexually abusing students and the outcomes of these cases, assuming of course that the federations would be willing to share these data.

This leads directly to the second obstacle, which concerned the federation affiliates' sensitivity to the issue and their resulting reluctance to share data. Discussions with teacher welfare or counseling personnel and senior officers of the federation affiliates revealed that in the past some of the affiliates had shared sensitive information related to this topic with journalists who later used the data in what the affiliates considered to be an irresponsible manner. Even more unfortunate, officers of one of the affiliates indicated that they had released sensitive information to an academic researcher who then used the data selectively and, in the affiliate's opinion, inaccurately to support an argument he or she was proposing. Unfortunately, two of the five affiliates declined to participate in the study. Although the reasons were not explicitly articulated, it is probably reasonable to assume that these two affiliates were concerned that their willingness to share data might be exploited yet again. Fortunately, the other three affiliates did agree to participate.

Regrettably, the noninvolvement of two of the five affiliates eliminated for all practical purposes the establishment of baseline data concerning accusations of sexual abuse by teachers in the Ontario secondary school and the Ontario Catholic separate school contexts. On the other hand, the participation of the other three affiliates provided a fairly comprehensive data set for Ontario's public (i.e., publicly supported, non-Catholic) elementary schools.

Participating affiliates were asked to complete a data form (see Appendix A) designed to gather data related to five recent school years. The form provided a format in which the affiliates could report the numbers of accusations,[4] numbers of members against whom charges had been laid, legal outcomes, and disciplinary actions taken by federations, school boards, and the Ministry of Education.[5]

Study 1: Results

Two incidents involving secondary teachers were reported by one participating affiliate. Because other participating affiliates did not provide data concerning secondary teachers, these two incidents represented anomalies and, therefore, were not considered in the analysis. Membership numbers for the affiliate in question were adjusted so as to represent only elementary teachers. In addition, one of the affiliates did not report data for the 1987-1988 and 1988-1989 school years. This was taken into account in relevant calculations. Table 1 provides a school year by school year summary of the data reported by the participating affiliates.

As the data in Table 1 illustrate, there were 47 incidents in which teachers belonging to these affiliates were charged with sexually abusing students during the five school years for which data were reported. Only one of the 47 cases involved a female teacher. In two of the cases charges were withdrawn before the issue could be determined by the court. In seven of the remaining 45 cases, the outcome was yet to be determined at the time the data were reported. Of the remaining 38 cases 22 resulted in the accused teacher being acquitted and 16 resulted in the accused teacher being convicted of the charges.

In order to establish some sense of the frequency with which teachers were charged with sexual abuse of students during the years surveyed by the study, Table 2 relates the numbers of teachers charged to the numbers in the total population represented.

The ratios ranged from .15 teachers charged per 1,000 in 1991-1992 to .20 per 1,000 in 1990-1991. The mean ratio for all teachers over the five-year period was .17 teachers charged per 1,000 teachers.

Obviously, because male teachers represented nearly 98% of those charged, it can be argued that the ratios cited above are misleading. For this reason ratios of numbers charged to numbers of male teachers are also included in Table 2. If male teachers only are considered, both in terms of the total number of teachers and the numbers of teachers charged, the ratios ranged from .51 charged per 1,000 male teachers in 1991-1992 to .77 per 1,000 in 1990-1991. The mean ratio over the five-year period for males only was .61 male teachers charged per 1,000 male teachers.

Study 1: Discussion

It must be noted that although it is clear that the overwhelming majority of charges of sexual abuse are leveled against male teachers, it is far too simplistic to assume that gender-related conclusions can be drawn from these numbers. Male teachers are clearly more vulnerable to problems created by traditional social stereotypes. For example, behavior that is interpreted as nurturing and desirable (e.g., hugging a student who is upset, assisting a child into or out of a snowsuit, etc.) when a female teacher is involved can easily be misinterpreted as having a sexual intent when the teacher is male.

Table 1
School Year by School Year Summary of Data Provided by Participating Federation Affiliates

School Year	*Total Number of Elementary Teachers in Affiliates Reporting*	*Number of Teachers Charged With Sexual Abuse*	*Sex of Teachers Charged*	*Legal Outcomes*	*Board Action*	*Ministry Action*
1987-1988[1]	47,639	8	8 Male	2 Acquitted 5 Convicted 1 Ongoing	2 Reinstated 1 Resigned 4 Terminated 1 No Action	5 Cancel Certificate 3 No Action
1988-1989[1]	50,045	9	9 Male	5 Acquitted 4 Convicted	4 Reinstated 3 Terminated 1 No Action 1 On Hold	2 Cancel Certificate 5 No Action 2 On Hold
1989-1990	57,050	9	9 Male	7 Acquitted 2 Convicted	5 Reinstated 1 Retired 1 Terminated 2 No Action	2 No Action 1 Not Reported
1990-1991	59,533	12	12 Male	6 Acquitted 1 Withdrawn 3 Convicted 2 Ongoing	6 Reinstated 1 Resigned 1 Retired 1 Suspended 1 Terminated 2 On Hold	1 Cancel Certificate 2 On Hold 9 No Action
1991-1992	61,299	9	8 Male 1 Female	2 Acquitted 1 Withdrawn 2 Convicted 4 Ongoing	3 Reinstated 1 Terminated 5 On Hold	6 On Hold 3 No Action
Totals	282,335	47	46 Male 1 Female	22 Acquitted 2 Withdrawn 16 Convicted 7 Ongoing	20 Reinstated 2 Resigned 2 Retired 1 Suspended 10 Terminated 4 No Action 8 On Hold	8 Cancel Certificate 28 No Action 10 On Hold 1 Not Reported

[1]One of the affiliates did not report data for the 1987-1988 and 1988-1989 school years.
Source: Affiliates of the Ontario Teachers' Federation.

Table 2
Descriptive Statistics Derived from School Year by School Year Summary of Data Reported by Participating Federation Affiliates

School Year	*Numbers Charged per 1,000 Teachers*	*Numbers Charged Per 1,000 Male Teachers*	*Known Legal Outcomes Expressed as a Percentage of Total Charged*
1987-1988	.17	.56	25% Acquitted 62.5% Convicted 12.5% Ongoing
1988-1989	.18	.63	55% Acquitted 45% Convicted
1989-1990	.16	.59	77% Acquitted 33% Convicted
1990-1991	.20	.77	50% Acquitted 8.3% Withdrawn 25% Convicted 16.6% Ongoing
1991-1992	.15	.51	22% Acquitted 11% Withdrawn 22% Convicted 44% Ongoing
Totals	.17	.61	47.8% Acquitted 4.8% Withdrawn 38.1% Convicted 16.6% Ongoing

Source: Affiliates of the Ontario Teachers' Federation.

Two conclusions can be reasonably drawn from these data. First, when viewed in the context of the entire Ontario public elementary school enterprise, few teachers are charged with sexually abusing students. Second, if one assumes that in Ontario individuals are not charged with serious criminal offenses such as sexual assault unless attorneys representing the crown have a reasonably strong expectation that the accused will be convicted, the withdrawal/acquittal rate for teachers who are charged appears to be high (in the area of 60%).

Unfortunately, the data in Tables 1 and 2 are insufficient to support any other conclusions or hypotheses. It is clear that the determination of any significant patterns (e.g., in grade levels taught by teachers charged, changes in frequency of charges over time, etc.), will require longitudinal data from a larger data base.

Study 2: Data Provided by Experienced Principals and Vice-Principals

In order to address the problem identified as the first obstacle in the original research design (i.e., federations had no reliable records of the numbers of accusations because many of these were handled informally at the school level), a second data gathering strategy was developed. This design involved polling experienced school administrators attending the annual Ontario Public School Teachers' Federation (OPSTF) Leadership Academy. An instrument

was designed for this purpose and was piloted at the 1993 Academy. Improvements suggested by results of the pilot study were incorporated into the instrument (see Appendix B), which was then used to gather data at the 1994 Academy.

The 1994 OPSTF Leadership Academy was attended by 33 experienced elementary vice-principals and principals representing 22 Ontario public school districts. Although OPSTF nominally represented male elementary teachers in Ontario's public schools,[6] roughly one fifth of the participants were female.[7] It should be noted that the group of principals and vice-principals attending the OPSTF Leadership Academy did not represent a random sample of the population of Ontario elementary school administrators. This group represented a convenience sample that was chosen specifically because its members were experienced elementary school administrators from a wide range of Ontario school districts.

These principals and vice-principals were asked to provide descriptive data concerning the schools for which they had been responsible for each of the past five complete school years (see Appendix B). In addition, each was asked to indicate if any teacher(s) on his or her staff had been accused of sexually abusing a student in any of these years. If any such incidents were identified, the administrator was asked to provide specific information concerning each such incident on a separate form (see Appendix C). These forms permitted respondents to provide detailed information concerning the accusation and its aftermath (i.e., charges laid, legal outcome, federation disciplinary action, school board disciplinary action, and Ministry of Education disciplinary action). Thirty-one respondents completed questionnaires.

Study 2: Results

Data supplied by these administrators provided an overview of 143 school-years.[8] Table 3 provides descriptive data concerning the grade levels of the elementary schools represented in the questionnaire responses.

Table 3
Grade Levels of Elementary School Years Represented in the Leadership Academy Sample

Grade Levels of Schools Represented in the Sample	*Number of School Years Represented*
Junior Kindergarten to grade 5	15
Junior Kindergarten to grade 6	16
Junior Kindergarten to grade 8	65
Kindergarten to grade 6	6
Kindergarten to grade 8	14
Grade 3 to grade 8	4
Grade 4 to grade 8	10
Grade 6 to grade 8	9
Grade 7 to grade 8	4
Total	143

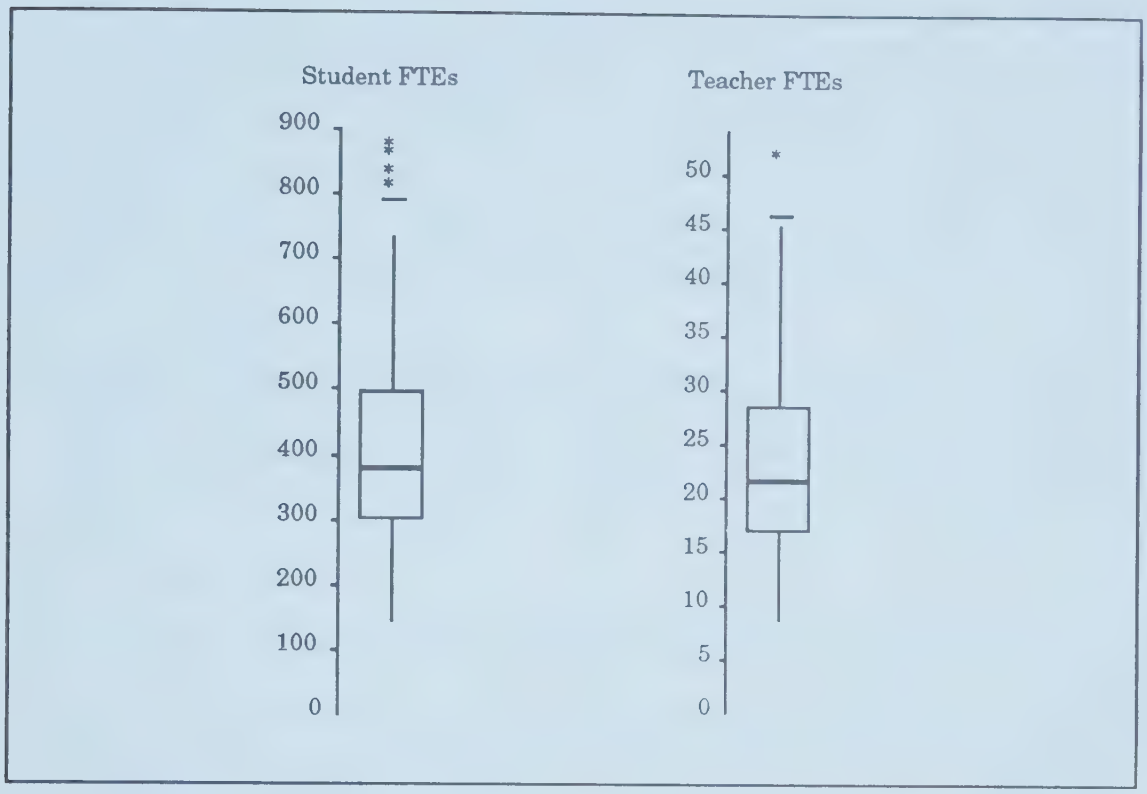


Figure 1. Box and whisker plots of teacher and student FTEs in schools represented in the Leadership Academy sample.

Staff sizes in the schools represented varied from a minimum of nine to a maximum of 52 with a mean of 24. Student populations varied from 145 to 863 with a mean of 405. Although the school sizes tended to represent a fairly normal distribution, most were clustered around the mean (as indicated by the box and whisker plots of teacher and student FTEs (full-time equivalents) in Figure 1. Only the largest of these elementary schools (student populations over 800) appear as outliers.

Table 4 contains the descriptive data provided by respondents by school year. Approximately 71% of the schools represented in the data reported for each school year were predominantly urban (i.e., at least 60% of the student body was urban), and 27% were primarily rural. Only 2% were described as having a student population evenly divided between students from urban and rural homes.

Obviously it was hoped that relationships would emerge that might link accusations of sexual abuse to school size, urban or rural environment, grade levels, and so forth, or that temporal patterns might become apparent (e.g., evidence that numbers of accusations, charges, or convictions are increasing or decreasing over time). However, no such relationships or patterns could be identified because in data related to 143 school years only four accusation incidents were reported. Table 5 lists by year the numbers of teachers accused and the ratio of the number accused to the number of teachers in the sample in each year.[9]

In two of the five academic years for which data were reported (1989-1990 and 1992-1993), none of the teachers in these schools was accused. During the 1991-1992 academic year two teachers were accused. One teacher was accused

Table 4
Descriptive Data Provided by Experienced Administrators in the Leadership Academy Sample, by School Year

School Year	Number of School Years Represented	Total Number of Teachers	Total Number of Students	Mainly Urban	Mainly Rural	Evenly Split Urban and Rural
1988-1989	24	545.5	9,561	17 (71%)	7 (29%)	0
1989-1990	28	676	11,313	19 (68%)	9 (32%)	0
1990-1991	29	696.8	11,471	21 (72.4%)	7 (24.1%)	1 (3.5%)
1991-1992	31	772.8	12,843	22 (71%)	8 (25.8%)	1 (3.2%)
1992-1993	31	736.8	12,692	22 (71%)	8 (25.8%)	1 (3.2%)
Totals	143	3,427.9	57,880	101 (70.6%)	39 (27.3%)	3 (2.1%)

in each of 1988-1989 and 1990-1991. Details of the incidents reported are outlined below.

1. During the 1988-1989 school year, a junior level (grades 4 to 6) student attending an average sized JK to 8 elementary school claimed a male teacher had touched her breast. The police were called and following their investigation advised that there was "insufficient evidence to charge" the teacher. Without implying the guilt of the teacher, the school administration, for the protection of everyone involved, modified the teacher's assignment in order to ensure that he would not be placed in a position where he would be required to work with small groups of students. No charges were laid and no disciplinary actions were taken.
2. During the 1990-1991 school year an 18-year-old senior level (grades 10 to OAC) student, obviously from a high school in the school system, claimed to have been sexually abused by a male teacher on the staff of a small, urban middle school (grades 6 to 8). Following administration interviews with the

Table 5
Numbers Accused and Ratio of Number Accused to Number of Teachers in the Sample, From Data Provided by Experienced Administrators in the Leadership Academy Sample, by School Year

School-Year	Numbers Accused	Numbers Accused Per 1,000 Teachers in the Sample	Numbers Charged
1988-1989	1	1.83	0
1989-1990	0	0	0
1990-1991	1	1.44	0
1991-1992	2	2.59	0
1992-1993	0	0	0
Totals	4	1.17	0

teacher and the student, the matter was dropped; no charges were laid and no disciplinary actions were taken.

3. During the 1991-1992 school year an intermediate level (grades 7 to 9) student attending a large, urban middle school claimed a male teacher was "continually coming around behind them [students] and touching their behind[s]." The student's parents were consulted and, after discussion with the school administration, agreed that the accusation was "unconfirmed." Once again, the matter was dropped; no charges were laid, no disciplinary actions were taken.
4. In the same year a junior level (grades 4 to 6) student attending an average size, predominantly urban school reported that a male teacher was involved in "inappropriate touching." The student later "admitted to making up the story." Obviously, no charges were laid and no disciplinary action was taken.

Study 2: Discussion

Only four incidents, none of which resulted in a charge being laid, were reported in the course of 143 recent school years. Clearly, although some basic ratios can be pointed out (e.g., on average, over the five year period, 1.17 teachers per 1000 were accused), no definitive relationships or patterns can be postulated on the basis of this data.

Summary Discussion

Although it is obviously impossible to draw many conclusions based on the results of the two investigations reported here, it is probably not inappropriate to suggest a few tentative observations that should be examined more closely.

1. Although it is self-evident that even one incident of sexual abuse by a teacher is one too many and represents a legitimate cause for concern and action, the data gathered thus far would suggest that the actual numbers of accusations, charges, and convictions are small given the numbers of individuals involved in the public school systems.
2. The acquittal rate for teachers appears to be far in excess of what common sense would suggest would be reasonable. The exceptionally high acquittal rate suggests problems may exist that are specific to this type of charge and/or to the teaching profession.[10] Further investigation is required.
3. In 1988 the Ontario Secondary School Teachers' Association reported that, among their membership, "spurious allegations of sexual assault.... have increased by 20 to 30 over the last five years" ("Teachers to fight," 1988, p. A4). The data from this study were insufficient to justify a similar conclusion relative to public elementary school teachers in Ontario. Analysis of a larger, more inclusive data set is needed. Nevertheless, it is clear that a number of teachers are falsely accused each year.

A larger, more inclusive research effort will be required before any more conclusive statements can be made concerning this phenomenon. However, the eventual success of such a research endeavor would not only be dependent on securing sufficient resources to do the job properly, but would also be directly proportional to the level of cooperation and commitment available from teachers' federations in all provinces. As the difficulties described in this study have shown, this level of cooperation is not always forthcoming.

Teachers Falsely Accused

It is not possible to determine from the data gathered in the studies described above precisely how many teachers are falsely accused of sexually abusing students. What is clear is that a significant number of false accusations occur each year. Given the serious personal and professional repercussions of such an accusation, this phenomenon clearly does represent a serious problem and also requires further investigation (Dolmage, 1995).

Conclusion

In March 1994 a professor (O'Neill) from the School of Organization and Management at Yale published an article in the *New York Times Magazine* that exploded the myth of the oft published and quoted lists comparing the most serious problems plaguing American schools in the 1940s (i.e., talking, chewing gum, making noise, running in the hall, getting out of turn in line, wearing improper clothing, and not putting paper in wastebaskets) with those of the 1980s (i.e., drug abuse, alcohol abuse, pregnancy, suicide, rape, robbery, and assault). Variations of these lists appeared on the CBS News, in *Harper's, Newsweek, The Wall Street Journal,* and *Time Magazine*; indeed, the *New York Times* published variations of the lists on five occasions. What O'Neill discovered was that these lists had no empirical basis whatsoever. Although the 1940s list may have been based on some relatively informal surveys of Texas teachers' opinions, the 1980s list was the creation of T. Cullen Davis, "a fundamentalist Christian who devised the lists as an attack on public schools" (O'Neill, 1994, p. 47). When asked for the source of his data, Davis replied, "How do I know what the offenses in the schools were in 1940? I was there. How do I know what they are now? I read the newspapers" (p. 48).

These lists were cited as data to support arguments at the highest levels of American educational policy making. They were quoted by "senators, mayors, state education officials, university professors, deans," a Surgeon General nominee, and even a former Secretary of Education (O'Neill, 1994, p. 48). Nevertheless, even though they had no basis in research, no one until O'Neill questioned their veracity.

What is most worrisome about the current discussions of teacher sexual abuse of students is that a folklore analogous to that cited above may be developing not only in the minds of the public, but also in the minds of Canadian educational commentators and policy makers who "know" the facts because, like Davis, they "read the newspapers." On the basis of anecdotal evidence the question of whether men should be encouraged to become primary teachers, or should be discouraged from working with young children at all, is being debated (Skelton, 1994). Intrusive searches of the backgrounds of prospective teachers have been recommended (Doctor, 1994). Perhaps this is only as it should be. There may be a serious problem in Canadian schools; there may be a need to institute intrusive teacher screening and supervision procedures in order to protect innocent children. On the other hand, there may also be a need to institute procedures to protect innocent teachers from character assassination. The point is we should know what we are talking about before we advocate significant public policy initiatives, particularly concerning significant and highly sensitive issues.

Defensible data would provide much more than a baseline from which to design further research to study the issues related to accusations of sexual abuse by teachers; such data could also be useful in more practical and immediate ways. Accurate data could, for example, provide both preservice and inservice teacher educators with more persuasive arguments concerning prudent practice for teachers facing the realities of the classroom in the '90s and beyond. Some hard facts might also encourage parents and trustees to adopt a reasoned, and perhaps less visceral, response to alleged cases of sexual abuse in the schools.

A credible source of data on the phenomenon should also force external agencies to adopt more responsible approaches to the problems associated with such accusations. The media, for example, might be persuaded to adopt a more conservative approach to the reporting of cases of alleged sexual abuse in the schools. Police officers and children's aid society workers might be forced to conduct thorough, comprehensive inquiries and to employ objective interview techniques in their investigations of cases of alleged sexual abuse in the schools. The courts might be urged to adopt a realistic, objective perspective when evaluating the reliability of evidence provided by child witnesses.

Should the data indicate that significant numbers of teachers are being charged with sexual abuse-related offenses and that a significant percentage of those charged are being found guilty, this would provide a powerful argument for the use of more stringent, and perhaps intrusive, screening procedures by teacher training institutions, certification agencies, and employing school boards.

On the other hand, should the data indicate that significant numbers of innocent teachers are being charged with sexual abuse-related offenses, such data would provide a powerful argument for the development of legal mechanisms to protect the personal and professional reputations of teachers who are falsely accused.

The problem is, of course, that at this point we have little more than opinions on which to base decisions concerning which of the above arguments we should pursue, if any.

Teachers' associations are clearly in the best position to provide the kinds of raw data that are needed if we are to develop meaningful responses to the fundamental questions this study has only begun to address; indeed, teachers' federations and associations probably already have the necessary records from which such data could be extracted. Teachers' organizations are not, however, ideally situated when it comes to the actual collection, synthesis, and analysis of these data, or the dissemination of the findings that result. Actual objectivity notwithstanding, they will always be seen as having a vested interest in the protection of their members and of the profession. Ideally, if the findings of such research are to be viewed as credible, this project will have to be a comprehensive, cooperative effort between academics and teachers' organizations. In the end, whether this can be done may depend entirely on the level of trust that can be established between these two groups.

The alternative is to allow the media to define the "reality" of this phenomenon for the public and for educational policy makers. Unfortunately, the media for the most part are simply businesses with advertising to sell and

shareholders to satisfy. They are inevitably attracted to the sensational and uninterested in the routine. As Ainsworth (1994) put it, "Good things happen in schools every day. It's not news" (p. 19); such a perspective cannot help but create a distorted view of what is really happening in Canadian schools. Policy based on such a distortion is unlikely to provide protection for innocent students or innocent teachers.

Notes

1. I have addressed the issues relating to false accusations in some detail elsewhere (Dolmage, 1995).
2. It is interesting to note that schools were the largest single source of referrals relating to child abuse and/or neglect and that school referrals had a higher than average substantiation rate. (Trocmé et al., 1994, pp. 102-103).
3. The Ontario Teachers' Federation is an umbrella organization that is composed of five largely independent affiliates.
 a. The Federation of Women Teachers' Associations of Ontario (FWTAO, 1993-1994 membership of 41,800) that traditionally (see note 7 below) represents female elementary teachers in the public (i.e., non-Catholic) schools;
 b. The Ontario Secondary School Teachers' Federation (OSSTF, 1993-1994 membership of 38,900) that represents secondary teachers in the public (i.e., non-Catholic) secondary schools;
 c. The Ontario Public School Teachers' Federation (OPSTF, 1993-1994 membership of 14,500) that traditionally (see note 7 below) represents male elementary teachers in the public (i.e., non-Catholic) schools;
 d. Association des enseignantes et enseignants franco-ontariens (AEFO, 1993-1994 membership of 6,700) that represents teachers in French first-language schools; and
 e. Ontario English Catholic Teachers' Association (OECTA, 1993-1994 membership of 32,200) that represents teachers in Catholic English first-language schools.
4. This proved to be an inappropriate question; see discussion above.
5. Now the Ministry of Education and Training.
6. Women could be associate members of the OPSTF, even prior to the Tomen decision (see note 7).
7. On March 31, 1994 the Ontario Human Rights Commission found that the compulsory allocation of teachers to particular affiliates on the basis of sex violated the Ontario Human Rights Code (the Tomen case). It would appear, therefore, that gender based federation affiliates will no longer be permitted in Ontario.
8. Ideally, data describing 155 school-years would have resulted (31 x 5). However, several of the administrators had worked in district central offices during the period. In addition, several had worked in secondary schools during one or more of these years; data from these years were not included in this analysis.
9. When making comparisons between the finding of Study 1 and Study 2, it is important to remember that in Study 1 federation affiliates reported numbers of teachers charged, whereas in Study 2 administrators reported numbers of teachers accused.
10. It would, of course, be helpful, and would give this point added relevance, if we could compare the acquittal rate for teachers with that of persons in other professions, or even with the acquittal rate for the overall population of persons charged with similar offenses. Unfortunately, such data are not readily available.

References

Abuse: An agonizing issue: Every teacher is vulnerable. (1990). *New Brunswick Teachers' Association News, 33*(2), 12.

Ainsworth, L. (1994). Managing the media: Some tips for teachers. *OPSTF News, 9*(1), p. 19.

Byrd, K.R. (1994). The narrative reconstructions of incest survivors. *American Psychologist, 49*(5), 439-440.

Cleared in sex case, teacher may quit job. (1987, December 27). *Globe and Mail*, p. A8

Cooney, H., Gardhouse, J., & Getty W. (1988). Coping with reality: Protecting teachers' rights. *OPSTF News, 2*(4), p. 21.

Criminal Code of Canada. (1985). R.S.C. c. C-46.
DeMitchell, T. (1981). What exactly is sexual misconduct? *Executive Educator, 168*(10), 33.
Doctor, E. (1994, June). *Sexual misconduct: Protecting students, teachers and school boards.* Paper presented to the annual conference of the Canadian Association for the Practical Study of Law and Education, Saskatoon.
Dolmage, W.R. (1995). False accusations of teacher sexual abuse of students: Exploring the dimensions of the problem. *Journal of Educational Administration and Foundations, 9*(2), 38-56.
Flavelle, D. (1988, January 9). Teachers vow to fight over "malicious" sex-assault charges. *Toronto Star,* p. B1
Gleaves, D.H. (1994). On "The reality of repressed memories." *American Psychologist, 49*(5), 440-441.
Gold, S.N., Hughes, D. & Hohnecker, L. (1994). Degrees of repression of sexual abuse memories. *American Psychologist, 49*(5), 441-442.
Kendall, D. (1990). Child sexual assault—It's time to talk. *OPSTF News, 4*(5), p. 32
Lesbian teacher accused of sex with students. (January 15, 1992). *Montreal Gazette,* p. A6.
Loftus, E. (1993). The reality of repressed memories. *American Psychologist, 48*(5), 518-537.
Loftus, E. (1994). The repressed memory controversy. *American Psychologist, 49*(5), 443-445.
Marczely, B. (1993). A legal update on sexual harassment in the public schools. *Clearing House, 66*(6), 329-31.
Natale, J. (1993). Kooks, crooks, and kids. *American School Board Journal, 180*(1), 18-23.
O'Connor, A. (1989). Seeking to protect those charged with sexual assault. *Reporter, 15*(1), 25.
Olio, K.A. (1994). Truth in memory. *American Psychologist, 49*(5), 442-443.
O'Neill, B. (March 6, 1994). The history of a hoax. *New York Times Magazine,* pp. 47-49.
Peterson, R.G.. (1994). Comment on Loftus. *American Psychologist, 49*(5), 433.
Polanyi, M. (1988, January 9). Union wants to charge girl, mother. *Globe and Mail,* p. A14.
Rabinovici, A. (1993, September). Medical experts explore legal issues arising from repressed memories of childhood sexual abuse. *Lawyers' Weekly,* p. 16.
Rabinowitz, D. (1993, May 1). A sex case built on air. *Globe and Mail,* p. D3.
Report of the Committee on Sexual Offenses Against Children and Youths. (1984). Ottawa: Canadian Government Publishing Centre.
Ross, V.J. (1985). Child molesting: Stop it before it starts in your schools. *Executive Educator, 7*(1), 22-24.
Shorten, L. (1994). False memory syndrome. *Canadian Lawyer, 18*(4), 16-20.
Skelton, C. (1994). Sex, male teachers and young children. *Gender and Education, 6*(1), 87-93.
Stewart, C. (1989a). False allegations: What is the real problem? *Reporter, 15*(1), 22-24.
Stewart, C. (1989b). A unique responsibility. *Reporter, 15*(1), 18-21.
Sundby, R. (1990). Guilty when proven innocent: Sexual allegations devastating. *Teacher, 2*(5), 7.
Teacher accused of molesting 11 girls in class. (1991, June 15). *Saturday Star,* p. A1.
Teachers to fight false sex charges. (1988, September 1). *London Free Press,* p. A4.
Teachers to sue false accusers. (1988, August 17). *London Free Press,* p. A4.
Teahen, K. (1988, January 13). Be more careful, abuse experts warn teachers. *London Free Press,* p. B8.
Toronto teacher cleared in sexual assault case. (1985, July 17). *London Free Press,* p. D7.
Trocmé, N., McPhee, D., Tam, K.K., & Hay, T. (1994). *Ontario incidence study of reported child abuse and neglect: Final report.* Toronto: Institute for the Prevention of Child Abuse.
Ulrich, T. (1989). Is it safe to be a teacher? *Manitoba Teacher, 68*(2), 9-10.
Wilson, D., & Littleson, C.M. (1987). Accused of molesting a student: Two true stories. *Learning, 16*(3), 42-48.
Yaworsky, W. (1992). Appropriate and inappropriate touch between teachers and students. *Canadian School Executive, 12*(4), 9-10.
Zakariya, S.B. (1988). How you can identify people who shouldn't work with kids. *Executive Educator, 10*(8), 17-21.

Federation Affiliate: _____ For School Year: 198___- 19_____ Page: ____

Teachers Accused of Sexually Abusing Students:

Ontario Teachers' Federation Affiliates Report Form

Case #	Teaching Level	Male/ Female	Interim Action by School Board	Charges Laid	Final Legal Outcome [Conviction/Acquittal/ Appeals]	Federation Disciplinary Action Taken	Final Action by School Board	Action by Ministry

For clarification or assistance in completing this data form contact:
Rod Dolmage, Faculty of Education, University of Western Ontario, London, Ontario, N6G 1G7

Phone: (519) 661-2087, Extension 8608

Appendix B
Questionnaire for Administrators

Principal ID Number: _____

School Identification

Because a given Principal may have been assigned to more than one school during the past five years and because the demographics of a particular school may change significantly from year to year, space has been provided for data relating to five schools.

1988-89 School School ID: 1

Grade Levels offered in the School (underline as appropriate):

JK K 1 2 3 4 5 6 7 8 9 10 11 12 OAC

Number of Teachers (FTEs): _____ Number of Students (FTEs): _____

Makeup of Student Body: % Urban.. _____ % Rural _____

1989-90 School [As Above (circle if appropriate)] School ID: 2

Grade Levels offered in the School (underline as appropriate):

JK K 1 2 3 4 5 6 7 8 9 10 11 12 OAC

Number of Teachers (FTEs): _____ Number of Students (FTEs): _____

Makeup of Student Body: % Urban.. _____ % Rural _____

1990-91 School [As Above (circle if appropriate)] School ID: 3

Grade Levels offered in the School (underline as appropriate):

JK K 1 2 3 4 5 6 7 8 9 10 11 12 OAC

Number of Teachers (FTEs): _____ Number of Students (FTEs): _____

Makeup of Student Body: % Urban.. _____ % Rural _____

1991-92 School [As Above (circle if appropriate)] School ID: 4

Grade Levels offered in the School (underline as appropriate):

JK K 1 2 3 4 5 6 7 8 9 10 11 12 OAC

Number of Teachers (FTEs): _____ Number of Students (FTEs) _____

Makeup of Student Body: % Urban.. _____ % Rural _____

1992-93 School [As Above (circle if appropriate)] School ID: 5

Grade Levels offered in the School (underline as appropriate):

JK K 1 2 3 4 5 6 7 8 9 10 11 12 OAC

Number of Teachers (FTEs): _____ Number of Students (FTEs) _____

Makeup of Student Body: % Urban.. _____ % Rural _____

Please indicate if any teacher or teachers on your staff was/were accused of sexually abusing a student or students in each of the past five school years.

School Year	Number of Teachers Accused
1988-1989	_____
1989-1990	_____
1990-1991	_____
1991-1992	_____
1992-1993	_____

Please fill out an "Accusation Data Form" for each accusation incident (a single form is appended: additional forms are available as required).

Appendix C
Accusation Data Form

ACCUSATION DATA FORM

Principal Identification Number: School Year: _____ Case # _____

Sex of Teacher (please circle): Male Female

Grade Level of the Student(s) (please circle): Primary Junior Intermediate Senior

Nature of the Accusation(s): _____

Interim Action By School Board (if any): _____

Charges Laid (if any): _____

Did the Accused Teacher Confess and/or Plead Guilty to the Charges (please circle)? Yes No

If Yes, Please Circle as Appropriate: Confessed Pleaded Guilty

Legal Outcome (please circle):

Conviction Conviction Upheld On Appeal Conviction Reversed On Appeal Conviction Under Appeal

Acquittal Acquittal Upheld On Appeal Acquittal Reversed On Appeal Acquittal Under Appeal

Comments Concerning the Legal Outcome (optional): _____

If Charges Were Not Laid, Please Explain Why This Was The Case: _____

Federation Disciplinary Action (if any): _____

Final Board Action (if any): _____

Ministry Disciplinary Action (if any): _____

Comments Concerning Federation Disciplinary Action, Final Board Action or Ministry Disciplinary Action (optional): _____

Any Additional Comments? (Please use back of sheet if additional space is required)

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 145-161*

K. Vern Stenlund
University of Windsor

Teacher Perceptions Across Cultures: The Impact of Students on Teacher Enthusiasm and Discouragement in a Cross-cultural Context

The purpose of this research was to document teacher responses relating to elements of their professional job enthusiasm and discouragement. Specifically, the study comprised secondary school teachers from seven nations involved in a cross-cultural research study initiated through the auspices of the Consortium for Cross-Cultural Research in Education (CCCRE), located at the University of Michigan, Ann Arbor. Teachers were first asked questions relating to sources of enthusiasm and discouragement in their professional lives. Through an inductive process their responses were categorized (clustered) into five major groupings, one of which was Students and Learning. Only data analyzed from this student cluster are reported in this article. Results indicate a consistency of response across all countries studied relative to the student and learning cluster. Specifically, respondents cite students as a primary factor associated with both professional enthusiasm and discouragement in an overwhelming statistical comparison. After responses were clustered into the five major headings, the student section accounted for over 50% of all responses associated with elements that create enthusiasm for teachers, although on the discouragement side 25% of responses were clustered in this category. In addition, among other major findings it appears that respondents do not consider student academic achievement as an important element in teacher enthusiasm across a majority of the cultures studied in this research. However, teachers do associate professional discouragement with students who exhibit low motivation while in the classroom. Finally, Japanese teacher responses are unique in two respects: first, they identify two-way communication problems as a prime source of discouragement; and second, the results appear to indicate that Japanese teachers believe that negative student behaviors and attitudes are less prevalent in their classrooms, resulting in lower scores in this category for the Japanese when compared with their counterparts from other countries.

Le but de cette recherche était de documenter les réponses des enseignants et des enseignantes en ce qui a trait aux éléments générateurs des sentiments d'enthousiasme et de découragement dans leur travail professionnel. Plus spécifiquement, cette étude a été basée sur des renseignements relevés des enseignants et des enseignantes d'écoles secondaires deuxième cycle de sept différentes nations qui ont participé à une recherche interculturelle initiée sous l'égide du Consortium des recherches interculturelles en pédagogie (CRIP)/Consortium for Cross-Cultural Research in Education (CCCRE) situé à l'Université de Michigan à Ann Arbor. Tout d'abord, on questionna les enseignants et les enseignantes au sujet des raisons qui engendraient l'enthousiasme et les sentiments de découragement dans leurs vies professionnelles. On a pu catégoriser et agglutiner les données inhérentes à leurs réactions par l'entremise d'un processus inductif selon cinq groupes majeurs, dont un sintitulait Élève et apprentissage. Cet article ne rapporte que les données analysées des élèves appartenant à ce dernier groupe. Les résultats indiquent qu'il y existe une consistance de réactions parmi tous les enseignants et enseignantes des pays étudiés en ce qui concerne les élèves et l'apprentissage. En raison des statistiques abondants provenants de l'étude comparative, les répondants et les répondantes citent plus spécifiquement les élèves comme étant un facteur primaire

Vern Stenlund is an assistant professor in the Faculty of Education. His recently completed doctoral dissertation at the University of Michigan (Ann Arbor) centered on job enthusiasm and discouragement specific to Canadian and Japanese teachers.

associé aux sentiments d'enthousiasme et de découragement professionnels. Après avoir rassemblé et d'agglutiné les réponses des enseignants et des enseignantes sous cinq en-têtes majeures, la section traitants des élèves comptait pour plus de 50% de toutes les réponses associées avec les causes qui engendraient l'enthousiasme chez les enseignants et les enseignantes tandis que le découragement ne représentait que 25% des réponses rassemblées dans cette catégorie. De plus, la recherche démontre qu'il semblerait que les répondants et les répondantes ne considèrent pas le succès académique des étudiants et des étudiantes comme étant un élément important qui engendrerait l'enthousiasme chez les enseignants et les enseignantes parmi la majorité des cultures étudiées. Cependant, les enseignants et les enseignantes ont certainement associé le découragement professionnel aux élèves qui démontrent peu de motivation en salle de classe. Enfin, au Japon, les réponses des enseignants et des enseignantes sont uniques de deux façons: D'abord, ils identifient les problèmes de communication à double voies entre l'enseignant(e) et l'élève comme étant une source primaire de découragement; et deuxièmement, les résultats semblent indiquer que les enseignants et les enseignantes du Japon croient que les comportements négatifs des élèves et leurs attitudes sont moins fréquents dans leurs salles de classe que dans d'autres pays. Ceci produit des scores plus bas dans cette catégorie pour le Japon comparé aux scores de leurs homologues de d'autres pays.

Introduction

The research as represented through this study is one portion of a larger project that compares sources of professional enthusiasm and discouragement for secondary school teachers in selected countries across a broad range of parameters. The study was conducted by the Michigan group associated with the Consortium for Cross-Cultural Research in Education (CCCRE). Presently 16 member nations are represented by the Consortium, of which seven have data submitted for this research including the United States of America (the State of Michigan, a four-county geographic range); England (the counties of South Yorkshire and Derbyshire); Germany (the state of Hessen); Japan (Chugoku and Kinki districts); Singapore (multiple districts throughout the city-state); Canada (school districts from the southwestern portion of Ontario); and Poland (sections of Warsaw proper). The CCCRE was initially established and based jointly in the University of Sheffield (England) and the University of Michigan (Ann Arbor) in 1978. Member nations have research teams located at various university centers in their respective countries and have been engaged in conducting research on common issues of teaching and schooling with small grants from government agencies, private foundations, as well as from their respective universities. The purpose of the Consortium is to generate basic behavioral science knowledge, applied knowledge about the nature of teaching and schooling, and policy and practice recommendations for the improvement of education within and between the cultural settings involved (Poppleton & Riseborough, 1989). The study reported here continues the research traditions of the CCCRE through reporting findings of a cross-cultural study involving the seven consortium member nations described above.

Purpose

Many individuals in the field of education encourage research that centers on various aspects of teaching practices, responsibilities, and climate that inevitably have an impact on student learning (Cross, 1987). Teacher perceptions with regard to students and student learning in a given teaching environment are one potentially fertile area of investigation as it relates to one's enthusiasm or discouragement regarding professional work. The teachers who par-

ticipated in this research have provided this type of information as a function of the purpose for the study.

To this end, the purpose for the research described here was to collect, compare, and analyze teacher perceptions specific to students and student learning vis-à-vis work enthusiasm and discouragement. In addition, the teacher perceptions investigated have been gathered in a cross-cultural context such that comparisons across and between cultures can be made.

Background

Enthusiasm and discouragement in school settings are important sources of influence on the work life of teachers. Collins (1976) and Cruikshank (1980) state that there is sufficient evidence of a strong, positive relationship between teacher enthusiasm and student achievement and that attentiveness to instruction increases when teachers presenting content possess enthusiasm. In comparison, discouragement has a strong negative influence on teachers and teaching. Blase (1986) points out that teacher stress is closely related to student discipline problems. Stress can become a source of discouragement for teachers that results in emotional and physical fatigue and a reduction in work motivation, involvement, and satisfaction. The larger CCCRE research project examined these issues within a comparative framework given a cross-cultural context.

It is important to describe the need for and validity of cross-cultural investigation and research in the educational community. As Judge (1988) states:

> Public perceptions of teachers and of their place in society vary between one country and another. A study of such variations is likely to generate new insights into social attitudes, not simply to teachers but also to education itself and to the values it embodies. (p. 143)

Kohn (1987) notes that without the benefit of cross-cultural or cross-national comparisons one cannot be certain that the information taken from single-nation studies is not simply the product of cultural or perhaps historical peculiarities specific to that particular country. This is an important concept and is particularly germane given the context of this study.

Recommendations generated through cross-cultural study are more apt to possess a broader, more global base versus similar findings extracted from a unicultural perspective. The value of cross-cultural research is summarized by Raivola (1985) who noted that this kind of study can prove invaluable as a frame of reference in comparative ventures, providing specific factors of education existing in one culture or environment that might not be prevalent in another. Some researchers believe that findings of similarities or differences evidenced in teachers between cultures can serve as catalysts for recommending effective change within a unicultural boundary (Purves, 1989; Triandis, 1978). They believe that there may be particular aspects of teaching (in this instance, in the realm of the secondary school teacher) that can be taken as universal and consistent around the world. Others such as Menlo and Poppleton (1990) argue that meaning without a cultural or social reference is incomplete, so information taken from such research out of context could be of little use. These poignant criticisms of the complexities that can arise from cross-cultural research should lead those involved in such studies to analyze

carefully how they choose to analyze and report their data before providing recommendations and/or commentary.

Methods

Organization of Data

Data used in this study were organized into five broad thematic cluster categories based on a system developed by the Michigan team affiliated with the CCCRE. These clusters were arrived at through thematic analysis done by an inductive process that allows for the creation of large categories (or clusters). Several studies to date have employed this categorization system with relative success (Evers, 1987; Marich, 1991; Menlo, Marich, Evers, & Fernandez, 1986). Responses were analyzed thematically by the five researchers involved in the collection of data and put into clusters that were distinctly alike. This process was initially conducted with each person separate, or blind, from the other researchers in order to determine what major cluster groups should be used. This analysis allowed for the development of several large categories. Once the researchers reassembled to discuss their individual thoughts and results, it became apparent that several major cluster groupings had evolved, including Students and Learning, Teachers and Teaching, Administrators and Administration, Work Conditions, and Parents and Community. These five clusters were eventually agreed on by the researchers as categories that would envelope the majority of responses generated through the study. The clusters (or cluster "trees") that were developed also consist of stems that act as major categories, with the attached branches serving as possible subcategory areas. Following this process, responses were reviewed in each major cluster and again sorted on the basis of alikeness in order to create the branches or subcategories. The five researchers read the responses and subsequently agreed on the classification of the responses for subcategorization. These subcategories were then assigned descriptive headings such as student enthusiasm and responsiveness, working and sharing with colleagues, and so forth. This classification procedure follows the principles outlined for the protection of emic distinctions in each culture, yet derives etic categories that accommodate distinctions (Brislin, Lonner, & Thorndike, 1973; Price-Williams, 1975). This implies that the teachers sampled in this study belong on a macro level to a broader group of people, namely, the world's secondary school teachers. This macro-world classification represents what Brislin (1986) called "the phenomenon, or aspects of it, which has common meanings across cultures under investigation" (p. 142). Countries can be compared on items that contribute to their shared (etic) aspects. However, when addressing differences in findings, careful attention should be given to possible historical, cultural, or political conditions not assessed that could skew or vary results. These differences might reflect distinct characteristics of a nation that might require more in-depth examination of the micro world that follows the emic tradition (Fernandez, 1991). Emic aspects of phenomena under investigation are said to be what is different and distinctive of each culture (Brislin, 1986). For the purposes of this study, only the information generated specific to Students and Learning is reported.

Participant Information

The teacher samples were self-selected as participants across all countries volunteered their time to engage in the interview protocol. Though voluntary by nature, this particular sample of secondary teachers does appear to reflect a broad cross-section of teachers in the educational environments investigated. And, although this methodology might not necessarily reflect a conventional construct in other research areas such as the natural sciences, it does allow for a practical and meaningful investigation of educator opinion and insight.

Interview Procedures and Data Collection

Many problems await the researcher who engages in research conducted in a foreign country. Besides the obvious considerations of language and local customs, there were also problems concerning available local assistance and the time frame provided for data collection. Given these types of issues a format was used that was deemed acceptable for all affected.

To appreciate more fully the nature of the research currently in progress (and the broad research questions investigated) a description is provided of the interview procedures and data gathering protocols unique to this study. A group inclusive, face-to-face format was employed where an open-ended process could be used. It was hoped that the information gathered would reflect reality as perceived and described by teachers for teachers, and a face-to-face protocol was deemed effective in this effort. Research questions were chosen over hypotheses to allow more flexibility on the part of respondents when providing answers. The resultant theme analysis is a common and broad technique often seen in the literature (Lortie, 1975).

Member countries associated with CCCRE research agreed on a consistent interview protocol from the outset of this study. Each group of researchers using this protocol attempted to conduct similarly structured interview sessions so that the teachers interviewed would experience a process as close to identical as possible. Interviews occurred with groups ranging from two to 16 teachers who participated in their own schools in a process that averaged approximately two hours in duration. Members of boards of education as well as school administrators cooperated with the research teams in scheduling interviews that took place during school hours. The teachers were initially given an explanation and description of the broad objectives for the research project and were then asked to consider their teaching experiences in the profession. A group facilitator (someone trained and familiar with questioning protocol) conducted semistructured, open-ended interviews. Each interview consisted of three main sections that all teachers were asked to participate in. During the initial portion of all interviews teachers were prompted to respond to the first broad question, namely, "What do you think are the things that generally 'turn teachers on' or could be described as sources of enthusiasm for teaching?" An assistant to the facilitator recorded all responses on a large piece of newsprint in full view of the participants. The information recorded was written verbatim with individual responses numbered for easy identification during computer coding at a later date. The facilitator was trained to probe for meaning and clarity if a teacher's response lacked specificity. Once the list was completed the facilitator then moved on to the second question and asked, "Looking at the list of responses generated, what two items from the list

personally 'turn you on' the most? Indicate the item number and description on the 5" x 7" index cards that are provided." Once this task was completed, the facilitator followed up with a third question asking why these two particular items selected "turned them on" so much. This concluded the first section of interview protocol.

The second part of the interview began with the corollary of the first set of questions. Specifically, teachers were asked what were the things that generally "turned teachers off" or could be described as sources of discouragement for teachers. The protocol was then repeated from the first interview section as information was gathered, clarified, and recorded on another sheet of newsprint. Having completed this task, the facilitator then asked teachers to again choose the two items that personally "turned them off" the most, and the teachers were similarly instructed to copy these responses on another index card as done in the first round of questions. Teachers were then asked as a final task to explain why these two items turned them off the most (question 3). This question finished the second section of the interview protocol.

In the third and concluding part of the interview, the facilitator pointed to all the answers generated on the various pieces of newsprint, both turn-ons and turn-offs. Teachers were then asked if there was anything that could be done to reduce the turn-offs (sources of discouragement) and/or increase the turn-ons (sources of enthusiasm). Again the facilitator probed all answers for clarity and the responses were recorded on a final separate sheet of newsprint. Teachers were then thanked for their participation and offered the opportunity to receive results of the study if they so desired. The interview was concluded and all data gathered (both newsprint sheets and index cards) were collected for coding and analysis. This third section of the interview data is currently being analyzed and is not included in this report.

It should be noted that during these interviews no direct scaling was attempted in any form, and no questions were asked specific to dimensions such as effort, locus, controllability, or difficulties. Rather, teachers were encouraged to produce their own lists consisting of whatever items they felt appropriate to their own experience in the profession of teaching. One of the benefits of analyzing this kind of data is that it attempts to avoid the biases that may be associated with research that constructs rigid categories and other classification parameters. These kinds of open-ended questions allowed teachers to provide broadly based responses across a variety of issues specific to the teaching profession. This freedom of expression was deemed a vital consideration by original members of the research consortium and is evidenced in various research protocols throughout the literature (Kottkamp, Provenzo, & Cohen, 1986; Lortie, 1975). Second, because of the language factor associated with any form of cross-cultural or cross-national research, researchers from across the consortium attempted to construct questions that would be universally understood and easily translated for practical use. Based on feedback from individual researchers through trial use in individual countries, the questions as used in this research were deemed to fulfil that criterion. They represent a simple yet effective means of deriving information germane to the topic of teacher enthusiasm and discouragement. However, I would be remiss if I did not discuss the issue of equivalence as it relates to response bias: for

example, equivalence in the sense that when accounting for possible differences we can be assured that the questions asked mean exactly the same things in each country (conceptual equivalence). Consortium members have tried to address this notion in the simplicity of construction and delivery of the questions utilized; equivalence in the sense that the people who answer the questions are the same kinds of people (sampling equivalence). All the people interviewed were secondary school teachers who volunteered their time for this research project and, as such, the researchers could not control for many factors that might influence responses across a volunteer or self-selected sample. However, if a bias across group responses in countries does indeed exist it is a consistent bias and one of circumstance, not creation. This should be remembered when discussing possible findings and conclusions emanating from this research. Finally, equivalence in that the methods of investigation and delivery of questions was as close to identical as possible (methodological equivalence). It should be stated as well that the Japanese data were collected in the presence of at least one American member of the CCCRE fluent in Japanese. To the extent that this represents continuity of collection protocols, the researchers were consistent vis-à-vis methodological equivalence.

Data Analysis

Frequency distributions were developed on the various patterns of similarities and differences between the responses from all countries involved. These nonparametric statistics are appropriate given the nominal type of data generated through this kind of study. The chi-square test for multiple independent samples (Siegel & Castellan, 1988) was used to determine if significant differences ($p<.05$) existed between responses of specific teacher groups across cultures with regard to sources of enthusiasm and discouragement.

It is important to consider the nature and context of this research undertaken given the methodology used. The educational community, as well as the process of education itself, does not easily fit into well-defined research normalacy and often requires unique ways in which reality can be reflected, if not measured. As a result, research in the boundaries of the educational environment often necessitates modes of inquiry that go beyond rows and columns of numbers. It must be so, lest our works succeed in doing nothing more "than reducing the richness of teaching to nothing more than the atomism of a multiple variable design" (Shulman, 1988, p. 4).

The data collected, reviewed, and analyzed through this research represent the combined efforts of research teams from many distinct countries and cultures. As members of the CCCRE, these researchers have cooperated in assembling, translating, and reporting the information herein contained. In reporting the findings, only major statistical categories are discussed to provide a general frame of reference in a comparative context. As well, lower- and upper-level interval percentages were generated to provide a conservative estimate of the confidence intervals. When upper- and lower-level intervals for any two countries overlap, they are not significantly different. If the intervals do not overlap, the country percentages are significantly different.

The data are presented anecdotally as well as in original table format. Original tables containing responses across all cluster findings are included to provide additional comparative data in areas such as percentage of response in

the specific cluster (students and learning) relative to the entire pool of responses. The findings generated through this work relate specifically to teacher perceptions surrounding students and learning as a function of their professional work lives.

All teacher responses were entered into a Macintosh microcomputer using Microsoft Word version 4.0. To delineate separate responses a master code was assigned indicating one of the following: country, school, number of groups, interviewer's code, date, and source category. This information was then entered into the MTS system at the University of Michigan. The coding was entered so as to allow for the generation of comparative tables.

Findings

The findings surrounding the nature of relations with students represents the dominant statistical category as witnessed in this particular study. Prior research has identified students as a primary source of teacher enthusiasm and discouragement (Menlo & Poppleton, 1990). This cluster focuses on such aspects as classroom practices, individual and group interaction with students, and the control one has over learning situations.

Researchers have determined that enthusiastic students positively influence the quality of work life for teachers (Collins, 1976; Cruikshank, 1980). Other researchers have stated that teachers are satisfied when their students learn well and actively engage in classroom activities with their teachers (Boland & Shelly, 1980; Kottkamp et al., 1986; Sarason, 1982). A corollary are those studies that note the impact that disruptive or uninterested students can have on teacher job dissatisfaction or discouragement (Greenberg, 1984; Litt & Turk, 1985). Still other researchers have postulated that the development and growth of students is a major source of job satisfaction for teachers (Dedrick, Hawkes, Richard, & Smith, 1981; Lobosco, Dianna, Sole, & Marciam, 1988), whereas other studies clearly imply that students are the focal point of job satisfaction and enthusiasm when they achieve academically and are enthusiastic and responsive (Evers & Engle, 1989). In light of the findings emanating from these and other studies, it is reasonable to suggest that teacher involvement with students can be characterized as a dynamic aspect inherent in the broad topic of teacher enthusiasm and discouragement.

Sources of Enthusiasm

The findings across the enthusiasm categories of this research support the literature described above. Of particular interest across all seven countries studied is the predominance of responses associated with enthusiasm generated through the Student and Learning cluster. As Table 1 illustrates, from the total response pool of 2,076 teachers issued 1,267 responses, or 61.03% of all statements recorded across the five cluster groups into the student and learning category. Participating teachers appear to associate students as the vital element that serves to enhance one's enthusiasm for the profession. Only the American and Singapore contingents fell below a 50% response rate with both countries close to the 50% level. The English and Japanese teachers both registered over 80% response rates in this particular cluster (80% and 82.47% respectively), a considerable result considering the context of the study. The specific subcategory Student Enthusiasm and Responsiveness (419 responses

Table 1
General Sources of Enthusiasm for Teachers in Seven Countries: Frequencies, Percentages, and Chi-square

Enthusiasm Categories	*Cross-culture Statistics* *Statistic Names*	*Cat. Sum*	*	*Within-culture Statistics* *USA*	*England*	*Germany*	*Japan*	*Singapore*	*Canada*	*Poland*
Students and Learning	Category Freq:	1,267	*	119	104	131	367	203	209	134
	Category Percent	61.03	*	46.48	80.00	61.79	82.47	47.21	56.49	57.51
	Upper % Limit	64.53	*	49.98	83.50	65.29	85.97	50.71	59.99	61.01
	Lower % Limit	57.53	*	42.99	76.50	58.29	78.97	43.71	52.99	54.01
Teachers and Teaching	Category Freq:	525	*	77	21	57	60	116	99	95
	Category Percent	25.29	*	30.08	16.15	26.89	13.48	26.98	26.76	40.77
	Upper % Limit	28.79	*	33.58	19.65	30.39	16.98	30.48	30.26	44.27
	Lower % Limit	21.79	*	26.58	12.66	23.39	9.98	23.48	23.26	37.27
Administrators and Administration	Category Freq:	121	*	29	0	2	8	50	32	0
	Category Percent	5.83	*	11.33	0.00	0.94	1.80	11.63	8.65	0.00
	Upper % Limit	9.33	*	14.83	3.50	4.44	5.30	15.13	12.15	3.50
	Lower % Limit	2.33	*	7.83	0.00	0.00	0.00	8.13	5.15	0.00
Work Conditions	Category Freq:	109	*	20	1	4	8	51	21	4
	Category Percent	5.25	*	7.81	0.77	1.89	1.80	11.86	5.68	1.72
	Upper % Limit	8.75	*	11.31	4.27	5.39	5.30	15.36	9.17	5.22
	Lower % Limit	1.75	*	4.31	0.00	0.00	0.00	8.36	2.18	0.00
Parents and Community	Category Freq:	54	*	11	4	18	2	10	9	0
	Category Percent	2.60	*	4.30	3.08	8.49	0.45	2.33	2.43	0.00
	Upper % Limit	6.10	*	7.80	6.58	11.99	3.95	5.82	5.93	3.50
	Lower % Limit	0.00	*	0.80	0.00	4.99	0.00	0.00	0.00	0.00
Enthusiasm Totals For all Countries	Category Freq:	2,076	*	256	130	212	445	430	370	233
	Category Percent	100.00	*	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 2
Students and Learning as Sources of Enthusiasm for Teachers in Seven Countries: Frequencies, Percentages, and Chi-square

Student and learning Subcategories	*Cross-culture statistics Statistic Names*	*Cat. Sum*	*	*Within-culture Statistics USA*	*England*	*Germany*	*Japan*	*Singapore*	*Canada*	*Poland*
Student enthusiasm and responsiveness	Category Freq:	419	*	32	46	35	83	102	93	28
	Category Percent	33.07	*	26.89	44.23	26.72	22.62	50.25	44.50	20.90
	Upper % Limit	36.58	*	30.40	47.74	30.23	26.13	53.76	48.01	24.41
	Lower % Limit	29.56	*	23.38	40.72	23.20	19.10	46.73	40.98	17.38
Emotional bond between students and teachers	Category Freq:	229	*	13	30	34	77	21	23	31
	Category Percent	18.07	*	10.92	28.85	25.95	20.98	10.34	11.00	23.13
	Upper % Limit	21.59	*	14.44	32.36	29.47	24.49	13.86	14.52	26.65
	Lower % Limit	14.56	*	7.41	25.33	22.44	17.47	6.83	7.49	19.62
Academic achievement of students	Category Freq:	132	*	23	14	18	22	31	7	17
	Category Percent	10.42	*	19.33	13.46	13.74	5.99	15.27	3.35	12.69
	Upper % Limit	13.93	*	22.84	16.98	17.25	9.51	18.78	6.86	16.20
	Lower % Limit	6.90	*	15.81	9.95	10.23	2.48	11.76	0.00	9.17
Growth and development of students	Category Freq:	239	*	22	3	32	81	13	45	43
	Category Percent	18.86	*	18.49	2.88	22.43	22.07	6.40	21.53	32.09
	Upper % Limit	22.38	*	22.00	6.40	27.94	25.58	9.92	25.04	35.60
	Lower % Limit	15.35	*	14.97	0.00	20.91	18.56	2.89	18.02	28.58
Long-term contacts with students	Category Freq:	67	*	17	1	0	33	4	6	6
	Category Percent	5.29	*	14.29	0.96	0.00	8.99	1.97	2.87	4.48
	Upper % Limit	8.80	*	17.80	4.48	3.51	12.51	5.48	6.38	7.99
	Lower % Limit	1.77	*	10.77	0.00	0.00	5.48	0.00	0.00	0.96
Respect appreciation from students	Category Freq:	69	*	7	9	0	4	24	17	8
	Category Percent	5.45	*	5.88	8.65	0.00	1.09	11.82	8.13	5.97
	Upper % Limit	8.96	*	9.40	12.17	3.51	4.60	15.34	11.65	9.48
	Lower % Limit	1.93	*	2.37	5.14	0.00	0.00	8.31	4.62	2.46

Table 2 (continued)

Student and learning Subcategories	*Cross-culture statistics Statistic Names*	*Cat. Sum*	*	*Within-culture Statistics USA*	*England*	*Germany*	*Japan*	*Singapore*	*Canada*	*Poland*
Students' understanding lesson and their role	Category Freq:	64	*	2	1	2	39	4	16	0
	Category Percent	5.05	*	1.68	0.96	1.53	10.63	1.97	7.66	0.00
	Upper % Limit	8.56	*	5.19	4.48	5.04	14.14	5.48	11.17	3.51
	Lower % Limit	1.54	*	0.00	0.00	0.00	7.11	0.00	4.14	0.00
When students come for help	Category Freq:	32	*	1	0	8	17	4	2	0
	Category Percent	2.53	*	0.84	0.00	6.11	4.63	1.97	0.96	0.00
	Upper % Limit	6.04	*	4.35	3.51	9.62	8.15	5.48	4.47	3.51
	Lower % Limit	0.00	*	0.00	0.00	2.59	1.12	0.00	0.00	0.00
Good student-student relationships	Category Freq:	16	*	2	0	2	11	0	0	1
	Category Percent	1.26	*	1.68	0.00	1.53	3.00	0.00	0.00	0.75
	Upper % Limit	4.78	*	5.19	3.51	5.04	6.51	3.51	3.51	4.26
	Lower % Limit	0.00	*	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Student and Learning Totals for all countries	Category Freq:	1,267	*	119	104	131	367	203	209	134
	Category Percent	100.00	*	100.00	100.00	100.00	100.00	100.00	100.00	100.00

or 33.07%) as seen in Table 2 shows this area as a key satisfier for teachers across all countries with the exception of Poland. The second highest scoring subcategory was Growth and Development of Students, with the seven countries providing 18.86% of all cluster responses in this area. Again, the English proved an exception along with the Singapore teachers as neither group expressed a substantial response pool in this subcategory. Filling out the top three subcategory scores on a percentage and frequency basis was Emotional Bond Between Students and Teachers. The results were similar to the previous subcategory with 229 responses (or 18.07%) associated in this subcluster. It would appear that teachers across the countries sampled identify with students personally vis-à-vis their own professional enthusiasm. Teachers consistently expressed a desire to have eager students who are responsive and attentive in order to increase their own professional enjoyment. As well, teachers appear to attach fairly significant importance to individual student growth and development and the bonds that develop between the teacher and student as precursory conditions through which the teacher gains enthusiasm for his or her work life.

One interesting note is that the subcategory Academic Achievement of Students placed a distant fourth in terms of total responses in this cluster. Only the American respondents had findings approaching 20% in this subcategory, the implication being that many of the teachers throughout this sample did not attach a high degree of importance to student achievement as it relates to their own professional enthusiasm.

Sources of Discouragement

In comparison with the response frequency for enthusiasm (1,267) teachers registered 565 responses (Table 3) out of a total response pool of 1,951 when commenting on students as sources of discouragement in their professional lives. Although only half the number of enthusiasm responses, teachers appear not only to identify students as the prime source of enthusiasm, but also as the prime source of teacher discouragement. Clearly teachers from most countries studied identified students as vital components of their professional lives in a positive and negative sense. On the discouragement side, teachers listed responses from the subcategory Low Motivation of Students as the single greatest discourager in the total of five subcategories examined with 36.81% of all responses falling into this area. As seen in Table 4, only the German teachers were out of the range relative to upper or lower limits, with a 22.09% response. Conversely, the Polish teachers scored the highest in this subcategory with a 60% response rate. It would appear that, in general, the teachers polled in this study agree that students who show low motivation are high on the list of elements that serve to discourage teachers in their teaching. This finding was closely followed by the second rated subcategory, Negative Attitudes and Behaviors, which accounted for 29.73% of total responses. In this instance, the Japanese were somewhat different in their feelings toward their students, as only 11.70% (or 20 responses) are associated with this subcluster. It would appear that of all the countries examined the Japanese have the fewest problems with negative attitudes and behavior problems associated with students as a discourager for teachers. Interestingly, although the Poles identified low motivation as a key problem area (60%) they did not feel the same about

Table 3
General Sources of Discouragement for Teachers in Seven Countries: Frequencies, Percentages, and Chi-square

Discouragement Categories	*Cross-culture statistics Statistic Names*	*Cat. Sum*	*	*Within-culture Statistics USA*	*England*	*Germany*	*Japan*	*Singapore*	*Canada*	*Poland*
Students and Learning	Category Freq:	565	*	43	37	86	171	78	110	40
	Category Percent	28.96	*	19.28	26.81	48.59	41.81	18.57	30.90	17.54
	Upper % Limit	32.55	*	22.88	30.40	52.18	45.40	22.16	34.49	21.14
	Lower % Limit	25.37	*	15.69	23.22	44.99	38.22	14.98	27.31	13.95
Teachers and Teaching	Category Freq:	319	*	34	28	20	114	48	55	20
	Category Percent	16.35	*	15.25	20.29	11.30	27.87	11.43	15.45	8.77
	Upper % Limit	19.94	*	18.84	23.88	14.89	31.47	15.02	19.04	12.36
	Lower % Limit	21.76	*	11.65	16.70	7.71	24.28	7.84	11.86	5.18
Administrators and Administration	Category Freq:	406	*	57	23	25	30	132	90	49
	Category Percent	20.81	*	25.56	16.67	14.12	7.33	31.43	25.28	21.49
	Upper % Limit	24.40	*	29.15	20.26	17.72	10.93	35.02	28.87	25.08
	Lower % Limit	17.22	*	21.97	13.07	10.53	3.74	27.84	21.69	17.90
Situational Work Conditions	Category Freq:	491	*	54	28	28	81	137	60	103
	Category Percent	25.17	*	24.22	20.29	15.82	19.80	32.62	16.85	45.18
	Upper % Limit	28.76	*	27.81	23.88	19.41	23.40	36.21	20.45	48.77
	Lower % Limit	21.57	*	20.62	16.70	12.23	16.21	29.03	13.26	41.58
Parents and Community	Category Freq:	170	*	35	22	18	13	25	41	16
	Category Percent	8.71	*	15.70	15.94	10.17	3.18	5.95	11.52	7.02
	Upper % Limit	12.31	*	19.29	19.53	13.76	6.77	9.55	15.11	10.61
	Lower % Limit	5.12	*	12.10	12.35	6.58	0.00	2.36	7.92	3.42
Discouragement Sums for all Countries	Category Freq:	1,951	*	223	138	177	409	420	356	228
	Category Percent	100.00	*	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Table 4
Students and Learning Sources of Discouragement for Teachers in Seven Countries: Frequencies, Percentages, and Chi-square

Student and Learning Subcategories	*Cross-culture statistics Statistic Names*	*Cat. Sum*	*	*Within-culture Statistics USA*	*England*	*Germany*	*Japan*	*Singapore*	*Canada*	*Poland*
Poor student-teacher communication	Category Freq:	106	*	2	7	18	68	1	5	5
	Category Percent	18.76	*	4.65	18.92	20.93	39.77	1.28	4.55	12.50
	Upper % Limit	25.86	*	11.75	26.01	28.03	46.86	8.38	11.64	19.60
	Lower % Limit	11.67	*	0.00	11.82	13.83	32.67	0.00	0.00	5.40
Negative attitudes and behaviors	Category Freq:	168	*	15	10	29	20	35	52	7
	Category Percent	29.73	*	34.88	27.03	33.72	11.70	44.87	47.27	17.50
	Upper % Limit	36.83	*	41.98	34.12	40.82	18.79	51.97	54.37	24.60
	Lower % Limit	22.64	*	27.79	19.93	26.63	4.60	37.78	40.18	10.40
Low achievement of students	Category Freq:	52	*	2	1	16	15	12	2	4
	Category Percent	9.20	*	4.65	2.70	18.60	8.77	15.38	1.82	10.00
	Upper % Limit	16.30	*	11.75	9.80	25.70	15.87	22.48	8.91	17.10
	Lower % Limit	2.11	*	0.00	0.00	11.51	1.68	8.29	0.00	2.90
Low Motivation in students	Category Freq:	208	*	18	17	19	56	26	48	24
	Category Percent	36.81	*	41.86	45.95	22.09	32.75	33.33	43.64	60.00
	Upper % Limit	43.91	*	48.96	53.04	29.19	39.84	40.43	50.73	67.10
	Lower % Limit	29.72	*	34.77	38.85	15.00	25.65	26.24	36.54	52.90
Students frustrated by the system	Category Freq:	31	*	6	2	4	12	4	3	0
	Category Percent	5.49	*	13.95	5.41	4.65	7.02	5.13	2.73	0.00
	Upper % Limit	12.58	*	21.05	12.50	11.75	14.11	12.22	9.82	7.10
	Lower % Limit	0.00	*	6.86	0.00	0.00	0.00	0.00	0.00	0.00
Students and Learning for countries	Category Freq:	565	*	43	37	86	171	78	110	40
	Category Freq:	100.00	*	100.00	100.00	100.00	100.00	100.00	100.00	100.00

negative attitudes or behaviors (17.50% of responses). Although one might surmise that low motivation could be easily linked to negative behaviors and attitudes, the Polish teachers appear to discount that connection.

Finally, Poor Student-Teacher Communications is the third rated discourager subcategory, primarily because of the response pattern from one country, Japan. The Japanese teachers are significantly different from all other countries studied in this regard, indicating that a lack of two-way communication often leads to teacher discouragement. It is worth noting that during interviews with some Japanese teachers a shared-blame problem often arose. This group of teachers appears to accept responsibility for failure in this regard far more often than virtually any other group of teachers, linking failure as much internally as externally. Indeed, the Japanese were the only group of teachers that had a higher frequency response in this subcategory than in the low motivation section. It should also be noted as well that as in the enthusiasm section, teachers from across the study did not associate professional discouragement with low student achievement. For the teachers studied in this research, student achievement did not play a vital role in either the enthusiasm or discouragement that teachers associated with their professional work lives.

Comments

In assessing the results of responses across all countries some interesting trends developed. As previously noted, teachers clearly identified students as the primary and central factor that has an impact on both their professional enthusiasm and discouragement. It would appear that teachers almost universally treasure student responsiveness and enthusiasm as a vital factor in their own enthusiasm, and conversely list low motivation in students as a discourager. The former point would confirm the findings of several studies previously referred to, including Collins (1976) and Cruikshank (1980) who concluded that student enthusiasm was one important aspect of teacher satisfaction. In addition, information derived from this study would also confirm the findings that Greenberg (1984) and Litt and Turk (1985) proposed in that disruptive or uninterested students can have a detrimental affect on teacher enthusiasm, leading to discouragement for educators. Based on the findings of this research, it appears that students are the vital bridge that links teacher enthusiasm and discouragement in a real time sense during the course of a teacher's professional engagement.

It is worth noting that teachers did not identify student success or academic achievement as vital to either enthusiasm or discouragement in their profession. This point could call into question some of the findings of Evers and Engle (1989) that tied teacher enthusiasm to student academic performance. Clearly the teachers from the countries sampled through this research indicated that academic success for students was not a key indicator for teacher enthusiasm or discouragement. Rather, it seems that it is the relationship between teacher and student(s) that counts far more than academic measures when discussing student impact on the professional lives of educators.

Finally, the results from this research continue to place the Japanese teacher in a different educational light. In a review of the comparative data from this seven-nation study, the Japanese stand out as the one group of teachers that appear different in many ways, especially relative to teachers and their com-

munication with students. Of particular interest is the notion of shared blame that the Japanese teachers often mentioned in the course of interviews. They appear to treasure two-way communication with pupils and readily accept their share of responsibility when problems arise between themselves and their students. Even the harshest critics of Japanese education must agree that this is a lesson in relationship building that teachers everywhere would be wise to study and emulate. The concepts of shared blame and two-way communication might explain in part the relatively lower responses of Japanese teachers with regard to negative student attitudes and poor behaviors. Undoubtedly all teachers, irrespective of their cultural affiliations, would welcome these circumstances in their own classrooms.

Summary

This research demonstrates that teachers want and need positive relationships with students in order to heighten their own sense of professional enthusiasm and worth. In addressing any potential changes or modification to school systems or curriculum, this vitally important notion should not be missed lest we rob teachers of the very reason they appear to become and stay enthused with their work. Teachers in this study repeatedly stated during the interview that students were at the heart of their existence in front of the classroom, and as members of the educational community at large we would be wise to acknowledge this fact during the construction of the new classrooms of the 21st century.

Through the type of cross-cultural research undertaken in this study educators might better be able first to understand, then effectively change, their professional environments for the better. This can only have positive ramifications for the classrooms of the future. As Emerson once wrote: "Nothing great was ever achieved without enthusiasm."

So it is with education.

Research Highlights

1. Teachers across cultures appear to identify students as the vital element that serves to enhance enthusiasm for teaching.
2. Teachers across cultures appear to identify students as the vital element that serves to enhance discouragement for teaching as well.
3. Student enthusiasm and responsiveness are an important aspect of teacher enthusiasm across a majority of cultures studied.
4. Student academic achievement does not appear to be an important element in teacher enthusiasm across a majority of cultures studied.
5. Students exhibiting low motivation appear to be a major contributor to teacher dissatisfaction across a majority of cultures studied.
6. Japanese teachers cite two-way communication problems as a major contributor to teacher discouragement.
7. Japanese teachers identify the fewest problems with negative student behaviors and attitudes of any nation studied.

References

Blase, J.J. (1986). A qualitative analysis of sources of teacher stress: Consequences for performance. *American Education Research Journal, 23*(1), 13-40.

Boland, P.A., & Selby, T.J. (1980). Factors associated with career change among secondary school teachers: A review of the literature. *Educational Research Quarterly, 5*(3), 13-25.

Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Lonner & J.W. Berry (Eds)., *Field methods in cross cultural research* (pp. 137-164). Beverly Hills, CA: Sage.

Brislin, R., Lonner, W.J., & Thorndike, R.M. (1973). *Cross-cultural research methods.* New York: Wiley.

Collins, M.L. (1976). *The effects of training for enthusiasm on the enthusiasm displayed by preservice elementary teachers.* Unpublished doctoral dissertation, Syracuse University.

Cross, P.K. (1987). The adventures of education in wonderland: Implementing education reform. *Phi Delta Kappan, 69,* 496-502.

Cruickshank, D.R. (1980). *Teaching is tough.* Englewood Cliffs, NJ: Prentice-Hall.

Dedrick, C.V., Hawkes S., Richard R., & Smith, J.K. (1981). Teacher stress: A descriptive study of the census. *NASSP Bulletin, 65*(445), 31-35.

Evers, T.B. (1987). *Factors affecting teacher job satisfaction in a number of high schools in Michigan.* Paper presented at the Annual Meeting of AERA, Washington, DC.

Evers, T., & Engle, J.M. (1989). *What teachers tell us about students and learning.* Paper presented at the Annual Meeting of AERA, San Francisco.

Fernandez, R. (1991). *Work attitudes among secondary school teachers in Japan and Michigan.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor.

Greenberg, S.F. (1984). *Stress and the teaching profession.* Baltimore, MD: Brookes.

Judge, H.G. (1988). Cross-national perceptions of teachers. *Comparative Education Review, 32*(2), 143-158.

Kohn, M.L. (1987, April). *Cross-national research in sociology: Presidential address.* Paper presented at the Annual Meeting of the American Sociological Association, Chicago.

Kottkamp, R.B., Provenzo, E.F., & Cohen, M.M. (1986). Stability and change in profession: Two decades of teacher attitudes, 1964-1984. *Phi Delta Kappan, 67,* 559-567.

Litt, M.O., & Turk, D.C. (1985). Sources of stress and dissatisfaction in experienced high school teachers. *Journal of Educational Research, 78*(3), 178-185.

Lobosco, A.E., Dianna L., Sole, K., & Marciam, E. (1988). *The intrinsic rewards of teaching: A book from the teacher's side of the desk.* Paper presented to the Annual Meeting of AERA, New Orleans.

Lortie, D.C. (1975). *School teacher: A sociological study in Chicago.* Chicago, IL: University of Chicago Press.

Marich, M. (1991). *What teachers tell us about teachers and teaching.* Paper presented at the Annual Meeting of AERA, Chicago.

Menlo, A., & Poppleton, P. (1990). A five-country study of work perceptions of secondary school teachers in England, USA, Japan, Singapore, and West Germany. *Comparative Education, 26,* 173-180.

Menlo, A., Marich, M., Evers, T., & Fernandez, R. (1986). *A cross-cultural comparison of the sources of professional enthusiasm and discouragement in teaching in England, West Germany, and the United States.* Paper presented at the Annual Meeting of AERA.

Poppleton, P., & Riseborough, G. (1989). *Teaching in the UK in the mid-1980s: The centrality of work in secondary teachers' lives in England.* Paper presented at the Annual Meeting of AERA, San Francisco.

Price-Williams, D.R. (1975). *Explorations in cross-cultural psychology.* San Francisco, CA: Chandler and Sharp.

Purves, A.C. (1989). *International comparisons in educational reform.* New York: Association for Supervision and Curriculum Development (ASCD).

Raivola, R. (1985). What is comparison? *Comparative Education Review, 29,* 3.

Sarason, S.B. (1982). *The culture of the school and the problem of change* (2nd ed.). Boston: Allyn and Bacon.

Shulman, L.S. (1988). Paradigms and research programs in the study of teaching: A contemporary perspective. In M.C. Wittrock (Ed.), *Handbook of teaching* (3rd ed., pp. 3-36). New York: Macmillan.

Siegel, S., & Castellan, Jr., N.J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill.

Triandis, H. (1978). Some universals of social behavior. *Personality and Social Psychology Bulletin, 4,* 1-16.

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 162-174*

Trevor J. Gambell
University of Saskatchewan

Ethnography as Veneration

Although ethnographic research in education is now an accepted and fairly established form of inquiry, ethnography presents problems for researchers who must position themselves throughout the process from initial planning through data collection, analysis, and reporting. This article suggests that the positioning of ethnographic researchers vis-à-vis their studies is sometimes precarious, leading to theoretical and linguistic difficulties that are often unresolved in the written reports. I suggest that one major problem is the tendency toward veneration of participants/informants and the environments they create, one of which is classrooms. This tendency, a tacit one, is often revealed in the discourse structures and language styles used by ethnographers when they write their studies. I discuss five concerns in the article, using examples from a variety of ethnographic studies in language education at various levels.

La recherche ethnographique dans le domaine de l'éducation est maintenant un moyen d'enquête plus ou moins établi et accepté. Malgré ceci, l'ethnographie présente des problèmes pour les chercheur(e)s qui doivent se situer pendant le déroulement de l'étude à partir de la planification initiale, à la collecte des données, à l'analyse, pour en arriver enfin au reportage de leurs trouvailles. Cet article suggère que la façon dont se situent consciemment les chercheur(e)s ethnographiques envers leurs études est parfois précaire, ce qui crée des difficultés théoriques et linguistiques qui demeurent souvent non résolus dans les rapports écrits. Je suggère que pour les chercheur(e)s la tendance de vénérer les participant(e)s et l'environnement qu'ils et qu'elles créent comme celui de la salle de classe est un des problèmes majeurs. Cette tendance, quoique sous-entendue et tacite, est souvent révélée entre les lignes des structures discursives et du style langagier qu'utilisent les ethnographes lorsqu'ils/elles écrivent et présentent leurs recherches. Dans cet article, je discute cinq points d'intérêt en utilisant des exemples tirés d'une variété d'études ethnographiques de l'éducation du langage à différents niveaux.

Introduction

Ethnographic research methodologies are commonplace in educational research of the 1990s. Especially in cases of graduate student research undertaken for theses and dissertations, ethnography is increasingly the methodology of choice. Increasingly too one finds ethnographic studies in published research journals and in academic books that report extensive studies. References such as Lincoln and Guba (1985) and Bogdan and Biklen (1982) are quoted to support the researcher's preference for ethnography as the research tool, be it participant-researcher study, case study, reflective study, interview study, or observational study. Along with ethnography, narration is a common mode of reporting, often intermixed with other modes of discourse such as description.

Also known as naturalistic study, ethnography offers researchers the opportunity to study educational phenomena in the settings where they occur,

Trevor J. Gambell is a professor and Graduate Chair in the Department of Curriculum Studies. His research interests are in the English language arts and the academic discourse processes of undergraduate students. He is writing an ethnographic study of the latter at present.

although settings do differ from the otherwise "natural" environment by the presence of the researcher or participant-observer. Ethnographers offer no apologies for their presence in the environment or culture; they make known their involvement and position themselves as a participant in the culture. They become themselves part of the research, part of the context, part of the phenomenon being studied. Through their becoming part of the context of study, in their immersion in the environment, by their interacting with research participants, ethnographer-researchers identify with the participants. In fact, identification probably begins before the formal outset or initiation of the study. Identification on a professional basis in many instances leads the potential researcher to the research, but that claim is the basis for a different article. However, if identification is not present before the study, it certainly develops during the study as the researcher seeks to establish a trustful relationship with (the) informant(s). The situatedness of researcher-participant among informants sets the stage for a possible, and even likely, bonding process wherein researchers develop empathy with the researched as they strive to make meaning of the events taking place and of the knowledge being created. But when researchers leave the environment, laden with data from a variety of sources, and begin the seemingly dispassionate process of creating meaning through analysis, the ethnographic bonding is likely to remain as a filter through which the data are analyzed and meaning or sense is made of the phenomena. Many researchers stay in contact with the environment through having participants read and comment on analysis and drafts. They believe that analysis is an ongoing process and guides data collection, also ongoing, in a reflexive way.

My concern in this article is that ethnographic bonding, or "going native" (Lincoln & Guba, 1985, p. 304), sometimes results in the human informants central to the study assuming a state of veneration as the findings are reported. The genesis of veneration is often first evident when the ethnographer-researcher describes the research setting, wherein a predilection toward creating a favorable impression of the "informant" is not always subtle. Reporters and investigators in the world of human affairs have always protected their informants because they realize that the information supplied by their informants and gathered in confidence must be taken seriously, accorded the status of truth or fact, or at the very least accepted as reliable evidence of the situation or phenomenon under investigation. Ethnographers, the investigators in educational research, likewise strive to establish the credibility of their informants. However, at times this research credentialing process borders on veneration of the principal informant(s).

Ethnomethodology

Heap (1992), an educational sociologist, describes ethnomethodology as a subdiscipline of sociology, although the disciplinary perspective has been used in linguistics and hence literacy studies in education. Atkinson (1990), a sociologist, argues that sociologists need to be self-consciously aware of what they are doing throughout the research process, and be more critically aware of how sociological texts are constructed by their authors. He hastens to add, however, that "attention to the 'literary' or 'rhetorical' features of ... texts in no way undermines their scholarly credibility or status" (p. 1). Atkinson's admonitions are as much applicable to educational researchers as to his own

fellow academics in sociology because the ascendancy of ethnomethodology in educational research has made such research, in effect, a type of sociological research.

Ethnomethodology examines the culture, society, or situation wherein the phenomenon of interest cannot be described or examined outside the context in which it occurs. The researcher is a participant or member of the culture or society under scrutiny and becomes part of the focus of the study as well as part of the data. However, the researcher as researched and as data brings about problems of its own, an issue to which I return in the discussion of Researcher as Tacit Informant. Atkinson (1990) addresses many of the problems inherent in ethnographic research from the sociological perspective, but educators have given their own color to ethnography.

For example, educational researchers have compressed the important aspect of time in ethnographic research; rather than extended time in the field, educational ethnographies often cover just a few weeks. Longitudinal studies in education are the exception. Another issue that comes up in educational applications of ethnography is that of the relationship between theory and research. One recent example comes from a proposal for doctoral research in adult literacy. The researcher-claimant drew heavily on grounded theory as the framework for the study. Quoting Strauss (1987), this writer declared that grounded theory is not a specific method or technique, but rather a style of doing qualitative research. The writer then declared that "the researcher does not begin with a theory, then set out to prove it. Rather, the researcher begins with an area of study and what is relevant to that area is allowed to emerge." I cannot imagine a researcher not approaching research without a theory, however tacit it be. No researcher is atheoretical, else there would be no identifiable study. The mere definition of a study, an area of study, presumes a theory, be it emerging or emergent. It behooves ethnographers claiming to be grounded theorists to proclaim and examine their tacit theory at the outset of their research and to return to it critically at the end of their study.

In an extensive article entitled "Some Similarities and Differences Among Phenomenological and Other Methods of Psychological Qualitative Research," Osborne (1994) hints at what may be a source of the difficulty of reporting findings faced by ethnographers, especially neophytes. This difficulty lies in the confusion between phenomenological research and ethnographic research. Quoting Spiegelberg, Osborne notes that phenomenological analysis "is analysis of the phenomena themselves, not of expressions that refer to them" (Spiegelberg, 1982, p. 690). Many ethnographers, however, seem to confuse the expressions that refer to the phenomena—that is, the language they use to describe the phenomena—with the phenomena. Is it possible to separate the phenomena from the language used to describe them? I believe it is, to the extent that one interprets one's linguistic descriptions rather than using and reporting them as raw data themselves, leaving them to the reader to interpret. In other words, the researcher needs to be critical of his or her own language of reporting. In this sense the ethnographer must be doubly reflective—of the phenomenon observed, and of the language of recording the phenomenon. Spiegelberg (1982) would concur: "Careful intuiting and faithful description

are not to be taken for granted and they require a considerable degree of aptitude, training and conscientious self-criticism" (p. 689).

My experience with graduate students writing theses and dissertations is that their ethnographic writing, especially in the narrative mode, exhibits a tendency toward venerating the situations and participants/informants under study. This tendency is not surprising, given that most educational researchers in graduate programs are teachers, are participating reflectively on their own and others' teaching practices, and are empathetic to the problems and issues facing teachers, including those participating in their own studies. The product of such veneration is an interpretive veneer that renders the data and findings as privileged understandings or knowledge. To illustrate how ethnographic veneration surreptitiously enters educational research, I have drawn studies from my area of interest, namely, English language arts education, language education, and the English language arts, from middle-years students to postsecondary adults. Examples are from master's degree theses, a doctoral research proposal, and research reported in journals and published in book form. Five concerns arising from ethnographic veneration are discussed: (a) The trope of the participant observer; (b) Ethnography as a product of research; (c) Ethnography as textual practice; (d) Researcher as tacit informant; and (e) Ethnography as construction of knowledge and social power. In the instances where I refer to studies undertaken at my own institution I use pseudonyms, not wishing unjustly to attack either the credibility of the researcher or the validity of the study.

The Trope of the Participant Observer

The first concern I wish to discuss is that of the trope of the participant observer, a term used by Herndl (1991) to deal with the researcher's stance or positioning of self in the study. (A trope is a phrase, sentence, or paragraph used to amplify or embellish the larger text into which it is inserted.) A dilemma exists for ethnographers when they build themselves into their own studies as both sources of data and as observers. They have to establish the uniqueness of their own experience through the I-was-there element, but they also have to suppress their own participation in the study in order to establish an unmediated or noninfluential authorization of the observation in order to render a degree of validity. Herndl (1991) suggests that this seemingly contradictory, dualistic act of presence and suppression is often accomplished by the use of the "arrival story, the poetic description of the ethnographer entering the native scene. This trope establishes the fieldworker's presence, authorizes her account, and then allows her to recede from the following description" (p. 325).

But alas, sometimes such poetic descriptions are less poetic than euphoric. An example from a master's thesis, a study of conferencing in a middle-years writing process classroom, sets this scene for the researcher's first visit to the classroom:

> One spring morning I made my first visit to P.'s classroom. The room was what I expected it to be, bright and well lit.... Hanging within the window and located throughout the entire classroom were numerous plants which provided a personal touch. The room had a pleasant, welcoming mood and I immediately felt comfortable.... Knowing there is a strong trend to improve the atmosphere of

> middle years classrooms, I recognized P.'s successful attempts to achieve this goal. (pp. 63-64)

I suggest that this writer has created a theoretical middle-years classroom setting into which he then steps and quickly confirms as the idyllic setting and which in turn matches the theoretical construct of a nurturing writing process classroom. The language used by the researcher, especially his choice of adjectives, suggests that an affirming stance has been taken; the writer has not suppressed his own participation in the study. Thus as reader, I am not convinced that a sufficient distancing has been achieved between researcher and researched. I develop a questioning, skeptical stance toward the validity of the study just unfolding.

The positioning of the researcher-participant in ethnographic research is critical. The researcher-writer can foreground herself or himself, as occurs in a study by Neilsen (1989) of the literate lives of three adults in Nova Scotia, and in the doctoral dissertation research of Cathro (1993) who studied the academic literate lives of two adult Aboriginal women in an undergraduate teacher education program; or position herself or himself in the background, as does Heath (1983) in her study of literacy among working-class white and black families in the Piedmont Carolinas. Either foregrounding or backgrounding can be successfully used by ethnographers, as these three studies exemplify, although foregrounding requires diligent use of descriptive language and analytic distancing that neophyte ethnographers may have difficulty achieving. The use of poetic description (trope) allows the writer to persuade the reader of the verisimilitude of the study.

Atkinson (1990) tells us that the text cannot simply transcribe or report. To be found plausible, the ethnographic text "must establish relations of identity and difference with other equivalent texts; it must establish relations of similarity and difference with the social world it reports" (p. 15). These other equivalent texts are those accepted ethnographic studies in the same field of study, accepted through the peer review process that leads to publication. Equivalent texts constitute the review of literature that informs the research at hand. However, some ethnographers are under the impression that their research is unique in the sense that nothing similar has preceded it in the ethnographic domain. Although this may be true, research from other research traditions or methods may have addressed the same phenomenon. Equivalent texts do not have to be of the same research pedigree or in the same discourse or rhetorical mode. Ethnographers need to situate their research within ethnography but equivalence is to the phenomenon of study; ethnographers need to read beyond ethnography. However, one's own study needs to be both compared to and contrasted with these other texts. It is simply not sufficient to affirm or privilege a dominant orthodoxy promoted by populist educational texts. I argue that this is what has occurred with ethnographic studies of the writing process, to the extent that the writing process has become an unchallenged theory that drives much uncritical and simplistic research in writing.

Ethnography as a Product of Research

The second problem is that of ethnography as a product of research. This occurs when the ethnographic text is a product of the ethnographer's dis-

course; the reader must be aware of how the ethnography was produced and how interpretation is, in part at least, a function of the writer's discourse. Herndl (1991) warns that the ethnographer's material is always a representation. For example, field notes "are already texts produced by the ethnographer's discourse" (p. 321). In this sense field notes are not raw data, as often claimed, but a linguistic representation of observed phenomena. Interpretation has occurred in the processing of recording. The "data" to which the ethnographer returns to write the ethnographic account is not an experience but the already textualized representation of experience in field notes and other records. Herndl argues that ethnography comes into being already inscribed in discourse. Except for verbatim transcripts of the researcher's interviews with informants, it is difficult to refute the interpretive nature of ethnographic data in the language of the researcher. Constructing the ethnographic account is a rhetorical activity; the critical ethnographic reader must be aware of the rhetorical nature of the text. How was it constructed and why? How were elements, "facts," data selected and why? How were such elements arranged in the text? Are there rhetorical patterns of comparison and contrast, projections of cause and effect, ends and means? These rhetorical questions need also be asked by the ethnographer in the process of constructing the text, although they are often neglected in the flow of the authoritative narrative voice.

Ethnography as Textual Practice

This concern is an outcome of the previous one. It is quite common to see the ethnographer invoke the narrative voice as the appropriate discourse for ethnography. This choice of discourse is an acceptable form of ethnographic writing, because it is the language of experience itself in the role of spectator (Britton, 1970). But narration must also be convincing as the authoritative voice of the ethnographer, and skillful use of narration that is also persuasive without being didactic does not come easily. Textually, the language employed must tell—there must be a sense of verisimilitude—and convincingly portray an ambience that readers inductively nuance. The latter requires careful use of descriptors: adjectives, verbs, adverbs. Here is an example of narration in an ethnographic study that calls into question textual practice.

> P.'s arrival was quick and cheerful. She took the time to say hello to a group of students arriving at the same time as she. It was apparent that her friendly informal nature was appreciated by the students. Too often teachers hide their personalities within the classroom and this, in turn, is recognized by the students. Not P. The energetic, bubbly P. that I knew was the same person welcoming the students to another day of school. (p. 67)

In this excerpt the writer's use of narrative is inconsistent. The use of the past tense suggests language in the role of spectator, removed from the here and now. But then the narrative quickly shifts in one sentence to a generalized, exculpatory discourse many times removed from the data, before returning to a laudatory sentence that reveres the informant.

Atkinson (1990) uses the term *reflexivity* to explain that texts do not simply and transparently report an independent order of reality. "Rather, the texts themselves are implicated in the work of reality-construction" (p. 7). The language of the ethnographer creates a reality that is a construct in and of itself,

and different from the reality of the lived experience, the actual data. The description of the teacher P. entering the classroom presents a different reality from the actual event.

The researcher-writer in the master's study quoted above describes the writing of his study as "an explanatory narrative." "In this study," he writes, "I used explanatory narrative to render an accurate account of my interpretation of conferencing in a middle years classroom" (p. 42). The writer cites Polkinghorne (1988) to support his use of this rhetorical technique: "The task of the researcher is to present ... patterns as realistically as possible. To accomplish this I used explanatory narrative as a framework. Narrative explanations are retrospective" (p. 43). Yet I charge that the language of this text is neither retrospective nor introspective. The explanatory discourse confirms the theoretical construct of an ideal middle-years classroom that preceded data collection. The phrase "the energetic, bubbly P. that I knew" suggests that the researcher had strongly identified with the informant prior to the study being conducted.

Once the researcher adopts the past tense, he or she adopts the role of spectator, reflecting and reporting on an experience or event and capturing that event not as a pure description of a phenomenon, but as a reflective account subject to interpretation that elapsed time and distance from the experience or event inevitably invokes. Narration in ethnography is not simply storytelling, but retelling. The researcher in the most recent example has confused the language of the here and now, language in the participant role, with the language of reflection, language in the spectator role.

There is no neutral language of observation. Researcher-writers need to be conscious of this fact and be sensitive to the language they employ, especially in their use of adjectivals and adverbials, the linguistic qualifiers. Atkinson (1990) writes of the "narrative contract" between ethnographer and reader via the rhetorical device known a "hypotyposis, the use of a highly graphic passage of descriptive writing, which portrays a scene or action in a vivid and arresting manner" (p. 71). I think that hypotyposis is what the quoted writer has attempted to use. The narrative contract is employed at key junctures in the text "to introduce settings and social actors, or to establish key transitions in the text. The reader enters into a sympathetic engagement with the social scene and its characters" (Atkinson, 1990, pp. 71-72). The last sentence is the key, I believe. The reader must be allowed to enter into a sympathetic engagement, but the writer should not predetermine the conditions for such engagement by taking on the role of the reader. This poses a dilemma for ethnographers using narration. The narrative mode invites the reader to become engaged with the text and context, yet academic discourse, the expectation many readers bring to an academic text such as a thesis, usually requires that certain conditions be met for positioning the reader. The writer becomes a reader of his or her text at a later stage but must use natural language at the point of narrative description. The writing process researcher cited above confuses the roles of writer and reader and empathetically engages himself through the explanatory narrative.

Narration, however, is a mode of discourse that developed outside the academy and is older than the academy itself. Narration is natural language; it is human for us to construct our world(s) through storying, such that one

telling or retelling of events is unlike another. Each recounting is a different story. Thus when we employ narration, a natural mode of discourse, in the service of ethnographic reporting, an academic and institutionalized discourse, and require it to take on other rhetorical functions such as persuasion, analysis, or cause and effect, we find that language breaks down, or our ability to use language for multifarious purposes cannot rise to the occasion. We fail or flounder as writers.

Researcher as Tacit Informant

I have foreshadowed the fourth concern in this discussion, that of the researcher as tacit informant in her own study. The unchecked or unrealized engagement of the writer as ethnographer discussed in the above section exemplifies this point. At what stage, if any, during the data analysis is the researcher able to separate self as data from informant data, and analysis of self-data from informant data? Or is this separation desirable or necessary? A theme throughout ethnography is the growth of the ethnographer that includes growth and understanding of self, often in a fairly explicit manner. This is often realized through the ethnographer locating herself or himself, and knowledge of self, in her or his study. Here is one example of the problem.

> Like many adult children of alcoholics, she has a driving need to achieve and exhibits the compulsive "bingeing" behavior that so often propels us through life. Like J. and me, E. pushes herself very hard. As she has grown, her activities have become highly constructive, and all have been carefully chosen to help her achieve her goals. (Neilsen, 1989, p. 120)

The same ethnographer comments on the research perspective of her study in these words:

> Working with J., J. and E., I was able to begin to see what literacy means in their lives. As I questioned and watched, I made connections between what I was seeing and hearing and what I had already experienced and known. What I learned from them not only gave life to the ideas about literacy and growth I brought to the research process, but extended them as well. (p. 7)

And finally, "In many ways, therefore, this work is about my literacy process as well" (p. 11). Earlier I noted that Neilsen (1989) had foregrounded herself in this study. It is not uncommon for the ethnographer to acknowledge his or her own growth, realization, broadened perspective, or self-revelation as a result of the ethnographic process. But it does beg the questions: Were the data interpreted solely or primarily through the unrealized researcher-self? Was the researcher as tacit informant the dominant voice in the study? Is the ethnography a study of writer-self embedded in the larger study, and which study is really being reported or foregrounded?

There is a danger of becoming underdistanced, too closely involved with one's topic or one's informants. Atkinson (1990) uses the term *marginal man* to describe the point of view of the sociologist ethnographer, one poised between intimacy and distance. "The participant side of participant-observation affords nearness, while the observer side lends fairness" (p. 19). Striking the right balance, finding the fulcrum, in this endeavor is a challenge for ethnographers. Lack of balance, which leads the researcher into adopting an advocacy role for the informant(s), comes through in the following excerpt.

> P. came over to me, smiled and said "hello" which immediately put me at ease. Before I had entered the room, I felt somewhat nervous but this quickly changed. Later, I said to P., "It's enjoyable being in the room, isn't it?" P. replied, "I like it. I really like it." (p. 67)

Clearly there is a tension faced by the ethnographic writer. "On the one hand there is the contrast between the 'self' of the observer and the 'other' of the observed.... In experiential terms the ethnographer is, in principle, always the 'marginal native'" (Atkinson, 1990, p. 157). The tension is between that of being a member of the community and being a stranger in the community. The temptation for the ethnographer, I believe, is wanting to be accepted as a member of the community, of desiring to be an insider rather than an outsider, when the ethnographer is not a member of the immediate or inner culture. However, in some instances the researcher is enculturated in the community, such as when a teacher-ethnographer researches events in her own classroom. And this encultured-ethnographer perspective presents other problems such as the valorizing of the teacher-informant; community members are mutually supportive and protective when necessary. Herzfeld (1983) expresses this tension in these words: "No ethnographer can ever claim to have been one or the other in an absolute sense. The very fact of negotiating one's status in the community precludes any such possibility" (p. 151). (The "one or the other" referred to by Herzfeld is insider or outsider.)

In the excerpt that follows, from a published study of the academic lives of two undergraduate students (Chiseri-Strater, 1991), it is unclear whether the writer is reporting what she learned in the class or what her informant heard or learned. The phenomenon of the researcher as tacit informant in her own study is evident in the depiction of an art history lecture in which the researcher sat along with students including her informant who was taking the class. Although the ethnographer scatters phrases such as "she comments," "she says" and "she suggests" when attributing statements to the observed instructor throughout the piece, the language is redolent of interpretation of content. The text reads like carefully transcribed lecture notes, in the guise of field notes.

> Hall (the art history instructor) often displays dissatisfaction with the slides, which cannot begin, she says, to do justice to the size or texture of the originals: "Oh nuts. This is a huge painting," she comments on Pollock's famous Autumn Rhythm: "Try to imagine this as filling up an entire wall of this room." The large scale of these works, she suggests, marks the final break of painting as being detached from the painter. In modern art, painting requires the viewer to be absorbed in pictorial space, the environment of the work encloses the spectator on all sides. As she moves closer to the slide, she suggests to students that they need to see the originals so that they can feel "the texture of the paint and allow themselves to float around in the painting." This kind of abstract art, Hall says, "requires you to enter into a dialogue with the painting itself." Again, generalizing from Pollock to an important concept of modernist work, she says that the abstract expressionist painters were "engaged in an argument between the literal surface and virtual space in painting." They were forging a new vocabulary for modern artists. (pp. 59-60)

This same study (Chiseri-Strater, 1991) yields another example of researcher as tacit informant. The researcher may, of course, claim that she is an informant

in her own study, but it must be made clear who the informant is at particular stages in the written research report, because interpretation is a singular act of the ethnographer in most cases. The researcher in the following excerpt is setting the scene for the art history lecture she audited. However, we are given an interpretive narrative wherein the scene is definitely a linguistic construct created by the researcher-as-informant and not the research informant, the student taking the course. Yet the subheading for this section of the book is "Anna in Avant-Garde Art Class." I believe it should read Researcher in Anna's Art History Class.

> I am early for art history class, which begins at four in the afternoon in Paul Arts, the building that houses the music, theater, and art departments. Waiting by the wooden weaving looms outside the lecture room, I observe the students milling around before class, some of them drinking coffee and tea purchased down the hall in the convenient art supply store. Eventually I join them with a cup of hot chocolate. Several students cluster together chatting softly, and although I cannot hear them I see by their dress that they are different: one bearded man in his late twenties has a scarf of rough South American fabric tied around his neck, a style that is seldom imitated on campus; another woman is wearing heavy work boots, splattered with paint, and olive clothing, which seems like a kind of military uniform. I glance down at someone's hands to see two-inch fingernails painted jet black, accompanied by an arm adorned with lovely clanging silver bracelets; when I look up I find hair that's partially dyed pink, gelled straight up from her head. (p. 57)

Ethnography as Construction of Knowledge and Social Power

Ethnography is an historically developed and institutionalized discourse. No matter that we claim narrative, or explanatory narrative, or poetic description as our preferred mode of reporting, ethnographic discourse is not in the domain of natural language or public language, except possibly through the writing of biography, but that genre usually falls into the broad category of realistic literature or literate discourse. Because ethnography has a history it is intertextual; it is derived from, is part of, and borrows from other texts in the academy. Ethnographic studies are situated (not exclusively) in other ethnographic studies, and this accumulation of cultural knowledge affirms ethnographic research findings as legitimate knowledge that in turn can be held up by the sites of ethnographic research (e.g., educational environments such as classrooms) as powerful agents with legitimized knowledge and power.

Because it constructs its own knowledge and empowers its informants through veneration, ethnography may also be the least effective means of critiquing contemporary ideology and practice. A result of ethnography's tendency to organize, construct, and present knowledge that is sympathetic and empathetic with its informants is that it loses its ability as an agent of educational change for those environments and individuals it studies. The ethnography of conferencing in a writing process classroom, one of the studies cited in this article, staunchly validates what it examines, or more rightly, describes and narrates. The ideology of the writing process is uncritically upheld through the veneration of its practice in one classroom. How can one develop a critical stance toward a topic wherein one or more informants have graciously

consented egress? The language of critical theorists is not the language of ethnographers. What we need are critical studies of ethnographies.

The Need for a Critical Approach to Ethnographic Discourse

Now that ethnographic research is becoming firmly established and accepted in education, we need to begin looking carefully at the practices and products of ethnography, including the preparedness of researchers to undertake ethnographic studies. I argue to this point that the ethnographer needs to be critically aware of his or her own use of language, of how language constructs knowledge and text that is different from the data derived from primary sources, including the verbatim words of informants. I suggest too that neophyte ethnographers, including graduate students, may not be fully aware of the rhetorical devices at work in ethnographic discourse and may lay naive claim to narration as the discourse of ethnography. Atkinson (1990) addresses these issues.

> The fully mature ethnography requires a reflexive awareness of its own writing, the possibilities and limits of its own language, and a principled exploration of its modes of representation. Not only do we need to cultivate a self-conscious construction of ethnographic texts, but also a readiness to read texts from a more "literary-critical" perspective. (p. 180)

In a particularly thoughtful and thorough naturalistic study of undergraduate students' writing and thinking in four disciplines, Walvoord and McCarthy (1990) worked from the point of view of what they call the "negotiated we" (p. 45). This technique involved their using coauthored drafts to achieve their aim of a research conversation. They openly acknowledge their concern about cooptation, or "going native." They write that "the danger existed that the outside investigator, Walvoord, might be so drawn into the world views of the discipline-based teachers that their interpretations would too much shape her own" (p. 45). Perhaps the "negotiated we" approach to ethnography, wherein an insider and outsider are collaborative researcher-participants as well as coauthors, presents a model for ethnography that minimizes the tendency toward veneration.

Walvoord's and McCarthy's (1990) method of tackling the textual problems of ethnography is similar to Clifford's (1983) proposal for a dialogic or polyphonic ethnography that includes a number of voices in order to disperse textual authority and to reflect the intersubjective, negotiated construction of ethnographic interpretation. Clifford borrowed the concept of multiple or dialogic voices from Bahktin's (1986) theory of the novel. A different approach has been proposed by Brodkey (1987) who argues for a narrative model of ethnographic knowledge. Her approach is akin to the strategy employed by Woolgar in *Laboratory Life* (Woolgar & Latour, 1979). Woolgar uses a fictional character to tell his story of the way scientists produce knowledge. The fictional character replaces the ethnographer in the scene of writing. Such a strategy would help to overcome narratives overlaid with other discourses as ethnographers struggle to narrate, describe, analyze, and explain in a singular rhetoric. Readers of ethnography may well be spared the likes of such memorable prose as:

> In her major field of study—art history—there's an undercurrent of resistance; the rebel is fighting the scholarship accommodator in an academic dance of virtual energies. As we go into her art class the following semester, we will remember this rebellious voice that is warning Anna of the dangers of the distanced stance toward art. Next frame please. (Chiseri-Strater, 1991, p. 57)

I mention above that biography may be approached as an ethnographic enterprise that employs narration, usually as a secondary genre, to create an engaging biographic account. A recent example of biographic writing illustrates this close relationship between narration and ethnographic biography. In his review of the book *Verdi: A Biography* by Phillips-Matz (1993), Rothstein (1994) tells the reader that

> Ms. Phillips-Matz's account, which is eminently plausible given her evidence, is unfolded with a calm detachment that serves her close devotion to her subject.... She ... has created a narrative that alters accepted understanding of matters both great and small. (p. 3).

It would seem that this biographer was able to use narrative to recreate and interpret the life of her subject, to whom she is devoted as biographer, through calmly detached narrative.

There seems to be no doubt that ethnographic research requires sophisticated research skills as does any systematic and serious method and process of inquiry. Knowledge of the theoretical perspectives underlying ethnography is essential, lest the naive ethnographer believe that ethnography is atheoretical or that theory derives from the study itself. It seems essential, then, that ethnographers be well trained and that they understand ethnography in relation to other qualitative research paradigms including phenomenology. I sense that some ethnographic researchers are confused between ethnography and phenomonology. Osborne (1994) suggests this is the case too.

> In phenomenological research ... there is an attempt on the part of the researcher to allow the data to speak for themselves in spite of the researcher's predispositions.... Although apparently more a priori in its approach, ethnographic research also stresses the importance of allowing the findings of the research to present themselves but through an inductive process. (p. 179)

If ethnography is a priori in its approach, it behooves ethnographers to make clear at the outset of their studies their relationship to the informants, the environment, and the phenomenon of study. As well, if the findings of ethnographic research present themselves to readers inductively, then the language of reporting data and the language of interpretation need to be distinctly different, especially if data are to speak of themselves rather than the researcher.

Ethnography has opened up possibilities for research that traditional research paradigms could hardly imagine, let alone cope with. Lest the reader misconstrue my message, ethnography as a research methodology is not the culprit. In short, my message is that ethnography, coupled with narration as the preferred discourse of reporting, makes analytic and linguistic demands on researchers that are far more rigorous and compelling than the predetermined and narrow constraints of traditional research methodologies. Ethnographers need to be not just solid researchers, but strong writers with a convincing

narrative voice and access to a vocabulary that is able to capture nuances of contexts and situations without usurping the interpretative function of readers.

Acknowledgments

I wish to thank Lorraine Cathro, Department of Curriculum Studies, University of Saskatchewan, and Mary Clare Courtland, School of Education, Lakehead University, who gave valuable advice through responses to the original paper. Thanks also to three anonymous reviewers of *AJER* for their helpful advice and specific suggestions incorporated in this article.

References

Atkinson, P. (1990). *The ethnographic imagination: Textual constructions of reality*. London and New York: Routledge.

Bahktin, M. (1986). *Speech genres and other late essays* (V. McGee, Trans., C. Emerson & M. Holquist, Eds.). Austin, TX: University of Texas Press.

Bogdan, R., & Biklen, S. (1982). *Qualitative research for education*. Boston, MA: Allyn and Bacon.

Britton, J. (1970). *Language and learning*. Coral Gables, FL: University of Miami Press.

Brodkey, L. (1987). Writing ethnographic narratives. *Written Communication, 4*(1), 25-50.

Cathro, L. (1993). *Many ways strong: Literacy, women, and academe*. Unpublished doctoral dissertation, University of Saskatchewan.

Chiseri-Strater, E. (1991). *Academic literacies: The public and private discourse of university students*. Portsmouth, NH: Boynton/Cook.

Clifford, J. (1983). On ethnographic authority. *Representations, 1*(2), 118-146.

Heap, J. (1992). Ethnomethodology and the possibility of a metaperspective on literacy research. In R. Beach, J. Green, M. Kamil, & T. Shanahan (Eds.), *Multidisciplinary perspectives on literacy research* (pp. 35-56). Urbana, IL: National Council of Teachers of English, and the National Conference on Research in English.

Heath, S.B. (1983). *Ways with words*. Cambridge: Cambridge University Press.

Herndl, C. (1991). Writing ethnography: Representation, rhetoric, and institutional practices. *College English, 53*(3), 320-332.

Herzfeld, M. (1983). Looking both ways: The ethnography in the text. *Semiotica, 46*(2-4), 151-166.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.

Neilsen, L. (1989). *Literacy and living: The literate lives of three adults*. Portsmouth, NH: Heinemann.

Osborne, J. (1994). Some similarities and differences among phenomenological and other methods of psychological qualitative research. *Canadian Psychology, 35*(2), 167-189.

Phillips-Matz, M.J. (1993). *Verdi: A biography*. New York: Oxford University Press.

Polkinghorne, D. (1988). *Narrative knowing and the human sciences*. New York: SUNY Press.

Rothstein, E. (1994, January 2). The furious man from Busseto. *New York Times Book Review*, pp. 3-4.

Spiegelberg, H. (1982). *The phenomenological movement*. The Hague: Martinus Nijhoff.

Strauss, A. (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Walvoord, B., & McCarthy, L. (1990). *Thinking and writing in college: A naturalistic study of students in four disciplines*. Urbana, IL: National Council of Teachers of English.

Woolgar, S., & Latour, B. (1979). *Laboratory life: The social construction of scientific facts*. Beverly Hills, CA: Sage.

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 175-187*

Michael W. Pratt

Tania Filipovich
Wilfrid Laurier University

and

Marie Bountrogianni
Hamilton Board of Education

Teachers' Views of Parents: Family Decision Making Styles and Teacher-Parent Agreement Regarding Homework Practices and Values

This study examined the possibility that teachers might endorse certain families' patterns in childrearing and in school-home coordination through homework more than those of others. A sample of 12 junior high school teachers and 36 parents of students from the same metropolitan school were interviewed regarding measures of decision making style, factors in student achievement, and the role of parents in helping with homework. Results showed that teachers endorsed an authoritative style of parent-child decision making, and were more congruent with authoritative parents on some aspects of attributions for children's achievement and on specific patterns of homework helping. However, all parents were more positive about their role in homework helping than were teachers. Possible implications of these patterns for teacher-parent interactions are discussed.

Cette étude examine qu'il est possible que les enseignants et les enseignantes puissent promouvoir les tendances et les orientations de certaines familles entre ce qui concerne élever les enfants et coordonner la communication entre le foyer et l'école par l'entremise des devoirs plus que les tendances présentes dans dautres familles. Un échantillon de 12 enseignants et enseignantes du niveau secondaire premier cycle et de 36 parents d'élèves de la même école métropolitaine ont été interrogés sur la façon de prendre des décisions, sur les facteurs qui contribuent au succès des élèves, et sur le rôle des parents qui aident l'enfant à faire ses devoirs. Les résultats indiquent que les enseignants et les enseignantes ont tendance à promouvoir le style autoritaire dans la relation parent-enfant pour les prises de décisions. Pour ce qui en est des devoirs, les résultats semblent être plus congruants avec les parents autoritaires en ce qui concerne certaines raisons attribuées au succès des élèves et certaines idées spécifiques en ce qui concerne aider les enfants avec leurs devoirs. Cependant, tous les parents étaient plus positifs dans leur rôle d'aider l'enfant avec ses devoirs que l'étaient les enseignants et les enseignantes. La possibilité de combiner ces idées pour permettre l'interaction entre les enseignants, les enseignantes, et les parents y est discutée.

Michael Pratt is a professor of psychology. He is a developmental psychologist with interests in social and cognitive growth and the family across the life span.
Tania Filipovich has an advanced degree in community psychology and is employed in the area of rehabilitation counseling.
Marie Bountrogianni is Chief Psychologist for the Hamilton Board of Education; her research interests are in home-school relations.

The two central settings for cognitive development in children's lives are the home and the school. How do the participants in these two settings view each other, and what factors influence their perceptions? Although there has been some theory and research on one influential individual difference factor in this context, cultural diversity, the present study focuses on another type of well-documented variability between families, parenting styles and practices (Darling & Steinberg, 1993). It addresses whether variations in such patterns may be related to teachers' views of parenting and school relations in the families of early adolescent students.

Baumrind (1973, 1991) reports a longitudinal study of the childrearing patterns of a sample of families from the preschool to early adolescent years. She describes three major parenting styles, based on two dimensions of parenting, warmth and support on the one hand, and demandingness or structure on the other. Authoritarian parenting is characterized by much structure and high expectations, but little warmth. In contrast, permissiveness as a style might or might not be characterized by warmth, but provides little in the way of structure or expectations of mature behavior. Authoritative parents combine an emphasis on structure and firm rule enforcement with warmth and an expectancy of mature behavior on the part of the child. At each age period studied, Baumrind found that authoritative parenting styles are associated with more social and cognitive competence in children.

This typology has engendered some recent research, particularly on adolescents' social and academic performance in homes characterized by these differing parenting patterns (Dornbusch, Ritter, Leiderman, Roberts, & Fraleigh, 1987; Steinberg, Elmen, & Mounts, 1989; Steinberg, Lamborn, Darling, Mounts, & Dornbusch, 1994). Authoritative styles were associated with the best school outcomes in all these studies, and this did not interact with ethnicity, suggesting that this parenting pattern is relatively advantaged across all groups studied (though, typically, stronger effects of authoritativeness are reported for Euro-American and Hispanic children than for other North American populations studied, Darling & Steinberg, 1993).

These studies of adolescents' personal qualities and outcomes have mainly focused on the child variables (e.g., child's work orientation) associated with the apparent success of authoritative styles in fostering school achievement. To date there has been only limited examination of the specific variations in parent values, beliefs, or behaviors regarding the school that might be linked to more authoritative styles and practices (Paulson, 1994). One important day-to-day link between home and school for older children is homework. Homework has been shown to play an important role in children's achievement (Keith & Page, 1985). In some recent reports it has been shown that variations in adults' parenting styles are associated with differences in homework tutoring patterns, such that more authoritative parents are more effective helpers (Pratt, Green, MacVicar, & Bountrogianni, 1992). In a related finding it has also been shown that parents who perceive themselves as more powerful causal agents with regard to children's school achievement are, plausibly, more likely than others to report assisting children with their homework (Pratt, 1993).

Thus parenting style, as defined by Baumrind (1973, 1991), seems to be clearly linked to children's achievement, as well as to some expectable varia-

tions in parents' views and practices regarding the home and school and especially their coordination through homework. However, a further important issue concerns teachers' views regarding parenting and the coordination of school and home. Some evidence has been gathered regarding teachers' views of parents of varying ethnic backgrounds (Chavkin, 1990). But there has been little study to date of teachers' views regarding appropriate styles and patterns of parenting and of parent cooperation and coordination with the school. The present study was designed to contribute some evidence on this point, as well as to compare teachers' views with variations in those reported by parents with different decision making styles.

A central question here concerns the matter of congruence between teachers and parents. Considerable argument, supported by quite striking observations, is that cultural divergence between teachers and families can interfere with children's schooling in various ways (Chavkin, 1990; Heath, 1983). For example, Heath (1983) showed that differences in linguistic styles and values across subcultures, and their congruence with the styles of teachers, played an important role in children's initial school adaptation. Others have similarly stressed the importance of linguistic dialect discrepancies between home and school, typically associated with ethnicity and class, which may hamper children's school adjustment (Meadows, 1993). One major element in such adjustment difficulty is undoubtedly the role of some teachers' negative attitudes toward such dialectical markers as signs of children's linguistic and cognitive (in)competence (Cazden, 1988).

It seems possible that teachers' values and ideas may be more or less discordant with those parents who exhibit distinctive parenting practices as well, although this question has apparently not been studied. It would seem a plausible hypothesis that an authoritative style of parenting and decision making might more closely resemble teachers' ideal for parents of adolescents in regard to both family process and school involvement. If this is so, it might be that a greater congruence of values between home and school in such families facilitates the coordination of these two central contexts for children's cognitive development. Thus this may be one aspect of the better school achievement associated with growing up in an authoritative family environment noted above. Of course, evidence of teachers' valuing of particular parenting styles could also raise the question of whether there is active discrimination toward children from families with nonoptimal parenting practices as well.

The present study tested the basic congruence question by comparing teachers' views about family and school relations with those of parents of young adolescents whose approaches were characterized by authoritative, authoritarian, or permissive decision making styles (Dornbusch et al., 1985). Thus the present investigation had two objectives: to describe teachers' views of ideal parent-adolescent relations regarding childrearing practices and schooling, and to compare these views with those of parents characterized by different types of family decision making styles.

Method

Participants

Participants in this study were 12 teachers and 36 parents (all but three were mothers) of students from the same urban junior high school in the metropolitan Toronto area. The participating teachers were recruited from the pool of grade 7 and 8 teachers at the school (approximately 30). Teachers were all born and educated in Canada.

The mothers were volunteers from three ethnic backgrounds, Anglo-Saxon, East Indian, and Greek, with equal numbers (12) in each group. In each of the cultural groups six participants were parents of boys and six were parents of girls. Mothers in the ethnic groups were typically first-generation immigrants to Canada who had attained at least a basic level of English and were able to communicate in the language. The children were enrolled in the grades 7 and 8 at the time of the interviews.

Tasks and Measures

Teachers and mothers completed a structured interview including a series of topics: parental decision making styles, ideas and practices regarding homework, and attributions for children's school achievement. Each of these topic areas is described below.

Parent decision making style measure. Our interview measure of family decision making was adapted from Dornbusch et al. (1985). Parents were asked about decision making practices regarding eight family issues, for example, choosing friends, clothes, spending money, watching television. For each item parents reported whether the issue was decided by: parents alone (authoritarian pattern), child alone (permissive pattern), or jointly by parent and child with discussion (authoritative pattern), following the procedures of Dornbusch et al. (1985). Decision making style was then characterized on the basis of the pattern of responses provided across the eight issues (each style could score from 0-8). Teachers were asked to report how a parent *should* handle decision making with a young adolescent around these same issues. Teachers' scores were obtained in the same manner as scores from parents.

General homework attitudes and practices. Teachers were asked five questions: if parents helped in their school, the frequency with which parents should help their children with homework, how parents feel when they are helping, the value that parents place on homework, and whether they personally have discussed homework issues with parents. Parents were asked the first four of these questions also, thus enabling direct comparisons between the parent and teacher groups. Cohen's kappas on 3-point scales for these items were all above .70.

Specific homework practices. Teachers and parents were asked about the specific homework practices that they used ("should use" for teachers) to help the child with mathematics homework, and about how they handled situations when the child was frustrated or refused to do homework. These three interview responses were content-coded for the number of mentions of various types of techniques and strategies. Nine categories were mentioned by 10% or more of the parents and teachers. An index of "good supervision" was created to summarize these patterns, weighting homework structuring and homework monitoring behaviors as positive, weighting giving the child the solutions,

threats, blaming the child and no response or no help as negative, and scoring mentions of discussion, referral to others, and direct tutoring as neutral (because the quality of these three behaviors could not be clearly specified from parents' oral reports). The correlation between two raters for scores on our summary tutoring index was $r(22) = .94$ for a sample of 24 transcripts.

Specific attributions for children's school achievement. Parents and teachers were both asked to attribute causes for children's school performance in six distinct academic areas: English, math, science, history/social studies, music or art, and physical education. The attributional choices included four categories, modeled after Hess, Chi-Mei, and McDevitt (1987): child ability, child effort, school/teacher factors, and home factors. The most important and next most important influences were selected for each school subject. The number of subject areas for which a cause was mentioned was obtained for both parents and teachers, so that scores could range from 0-6 for each cause.

Procedure

Parents were interviewed individually in their homes by a graduate student. Teachers were interviewed individually at school by the same student during nonclassroom time. The interviews took approximately 30-40 minutes to complete and included some additional topics regarding schooling not relevant to the present focus on homework.

Results and Discussion

Perceptions of Ideal Parenting Styles

Teachers and parents both responded to questions regarding decisions about eight family issues. Teachers were asked about the ideal pattern of family decision making for junior high students and parents, whereas parents reported on their own family style. Decision making style categorizations were established on the basis of the most frequently chosen pattern (of the three choices of child-alone, parent-alone, joint parent-child). This resulted in the classification of 14 parents as authoritative across the sample, 11 as permissive, and 11 as authoritarian. For teachers, 11 were classified as authoritative, whereas one was classified as authoritarian. Clearly the authoritative pattern is more common for teachers as an ideal than for parents in practice overall, chi-square = 8.04, (1, N=48), p<.01.

Table 1 shows the mean scores across the eight items for teachers and for the three parenting style groups in the sample. A Parent Decision Style/Teacher Group (4) MANOVA on the three patterns revealed a significant Group X Measure interaction effect, Pillai-Bartlett Trace $V(6,88)=19.43$, $p<.001$. There was also a significant measures effect, Pillai-Bartlett Trace $V(2,43) = 5.54$, $p<.01$, which reflected the fact that parent-only decisions were less common than all others across the sample (see Table 1).

Follow-up univariate analyses on each of the interview measures revealed that permissive parents were significantly more likely than all others to report child-only decisions, $F(3,44)=16.33$, $p<.01$. Scheffe tests at the .01 level confirmed that this group differed from teachers and from authoritarian and authoritative parents; these latter three groups did not differ from each other. Parent-only decisions were more likely to be reported by authoritarian parents, $F(3,44) = 20.78$, $p<.01$. Scheffe tests confirmed that this group differed from all

Table 1
Decision Making Profile Means by Teacher and Parent Style Groups, and by Parent Ethnic Group

	Decision Making Style		
	Child-Only	*Parent-Only*	*Joint*
Parent Style Group			
Authoritative Parents	2.50	1.86	3.64
Permissive Parents	4.64	1.91	1.45
Authoritarian Parents	2.18	4.09	1.73
Teachers	1.92	0.83	5.25
Unweighted Means	2.81	2.17	3.02
Parent Ethnic Group			
Anglo Parents	3.58	1.67	2.75
East Indian Parents	2.92	2.92	2.17
Greek Parents	2.67	3.08	2.25

others, and that none of the other groups differed from each other at $p<.01$. Finally, joint parent-child decisions were more common for both authoritative parents and for teachers, $F(3,44) = 26.75$, $p<.01$. Scheffe tests at the .01 level showed that teachers and authoritative parents did not differ from one another, whereas each of these groups differed from the authoritarian and permissive groups (see Table 1).

The results of these analyses are unsurprising, given that the parental decision making style groups were in fact established on the basis of this criterion measure. However, the analyses do reveal that teachers' ideals for parent-child decision making resembled the authoritative pattern most closely in stressing the importance of joint decision making in families of early adolescents. Inspection of teachers' responses by specific issue indicated that joint decision making was the ideal for all of the eight issues except for choosing clothes and after-school activities, for which child control was somewhat more commonly recommended.

For comparison purposes, means for the three ethnic groups of parents are also shown in Table 1. There were no differences between these three groups on any of the three decision making style measures in an overall MANOVA. Thus differences in parents' reports of practices are not strongly linked to ethnic variation in this urban junior high school sample, although it must be acknowledged that these samples are small and provide low power for tests of subcultural differences.

General Ideas Regarding Homework and Parenting

Teachers and parents were asked if parents helped in their school, how often parents should help, how parents feel when they are helping with homework, and how much parents value their children having homework. Only one of the 12 teachers reported that parents helped children with homework in any useful way. Some mentioned that parents' efforts probably hindered children:

> I think most parents get so emotional about it, they get so upset when their children don't understand something. They have very little patience and very little understanding about how to teach something, and the overall effect of their help is probably negative in most cases ... They're too emotional and as a result, they're very destructive in the process.

In comparison, more than half of the parents believed that other parents in the school helped their children with homework, and this did not vary by decision making style group substantially (see Table 2). Small cell sizes prevent a statistical analysis of overall parent versus teacher differences utilizing a chi-square statistic, but they are clear on inspection.

Both parents and teachers discussed how often parents should help, with the options being "on a regular basis," or "only when asked." The percentage of parents and teachers who felt that helping should take place on a regular basis is shown in Table 2.

Small cell sizes prevent an overall test of these differences, but teachers were clearly more skeptical of the amount of helping that parents should engage in than were parents themselves.

Both parents and teachers were asked how parents feel when they help their child with homework, rated on a 1-3 scale. The means for these ratings are shown in Table 2. As can be seen, teachers generally perceived that parents enjoyed this less than parents themselves actually reported. A Parent Decision Style Group/Teachers (4) one-way ANOVA on these ratings revealed only a borderline effect, $F(3,43) = 2.50$, $p=.07$. Follow-up Scheffe tests revealed no significant differences between groups at the .05 level. However, a contrast testing the differences between the three parents groups combined and teachers was significant, $t(45)=2.79$, $p<.01$. Many teachers' comments echoed one who said:

> I think most parents haven't got the education to help the kids, even at a low level, because they haven't covered the topics that are being covered today ... They don't really know the answers so I don't think they feel confident in helping.

When asked how much parents valued homework, most teachers (8 of 10 whose responses could be coded) reported that parents wanted children to

Table 2
General Attitude Means and Percentages Regarding Homework Helping by Parent-Teacher Groups

	Help Regularly	*Should Regularly*	*Enjoy Helping*[a]	*Homework Valuable*
Authoritative Parents	57%	72%	2.14	100%
Permissive Parents	64%	45%	2.18	100%
Authoritarian Parents	45%	91%	2.10	100%
Teachers	9%	36%	1.42	80%

[a]Rated on a 1-3 scale, "not at all" to "a lot."

have regular homework, although they were sometimes skeptical of parents' reasons:

> I think [parents value homework] too much in the wrong way. I think a lot of parents look at homework because they did a lot of homework, which was a lot of rote type of material ... I get the feeling that more often than not they want the traditional type of homework rather than the idea of projects, long-term assignments, which is what we are trying to do more of.

As can be seen in Table 2, all parents in all decision making style groups reported that they felt that homework had considerable value. Finally, about half of the teachers reported that they discussed homework with parents regularly (6 of 11), although most of the others indicated that they thought this was a good idea, and were considering doing it in the future.

Teachers and parents thus seemed to differ considerably in their perceptions of homework, with parents generally more positive about their role in helping with it and their enjoyment of that role. In general, this divergence between teachers and parents appeared to cut across the three parenting groups (see Table 2). There were no significant differences among parents in parallel analyses by ethnic group regarding these homework issues either. Teachers clearly seemed to have something of a negative stereotype about how parents and children experienced and used homework. However, a number of the teachers remarked that they would like to work more closely with parents on homework issues, and most did acknowledge the importance of homework in parents' values.

Specific Parental Help with Homework

Teachers and parents were asked more specifically how parents help ("should help" for teachers) their children with homework, and then how they would handle a situation where the child did not want to do homework, and one where the child was frustrated. The responses given were coded into nine categories, and then weighted and summed across the three interview items to provide an index of Good Helping, based on the principles of effective tutoring. This self-report index for parents correlated r=.42, p<.05, with observers' ratings of parents' actual tutoring in a separate homework session (Pratt, Hunsberger, Pancer, Roth, & Santolupo, 1993), providing some indication of its construct validity.

A Parent Decision Group/Teachers (4) ANOVA on this index indicated a significant effect, $F(3,43) = 8.49$, p<.01. Means were –.64, –.36, .43, and 1.45, for permissive, authoritarian, and authoritative parent groups, and for teachers respectively. Follow-up Scheffe tests revealed that teachers differed from permissive (p<.01) and authoritarian (p<.01) decision style parents, but not from authoritative parents. Means for a parallel Parent Ethnic Group/Teachers (4) ANOVA were –.66, .50, –.17, and 1.45 for Anglo, East Indian, and Greek parents and teachers respectively. Follow-up Scheffe contrasts on this analysis revealed significant differences between teachers and both Anglo and Greek parents at the .01 level.

Thus teachers were shown to endorse an effective tutoring style on this rationally constructed index; parents who used an authoritative, joint decision making style resembled them the most on this measure, compared with other

parent style groups. Almost all the teachers mentioned structuring the child's environment or time (83%), and many mentioned monitoring homework assignments and completion (43%), the two positive strategies in our index. In contrast, the negative strategies (giving solutions, threats, blaming the child, and no response/no help) were never mentioned as positive by teachers, and were explicitly proscribed by several of them. Discussion and direct tutoring were also mentioned fairly often by teachers (33% and 43%, respectively). Overall, then, teachers' responses regarding how parents should help closely resembled the ideal pattern established here for parent tutoring, providing support for this index of good homework helping. And teachers' prescriptions most closely resembled authoritative parents' reports of what they actually did in helping with homework.

Attribution Patterns for Student Performance

As noted in the introduction, a previous analysis of attributional patterns of parents in another investigation indicated that more extensive homework helping was consistently linked to a feeling of more responsibility for children's school achievement (Pratt, 1993). In an earlier report on the parents in this sample (Pratt & Sebastian, 1989), we indicated that there were no ethnic group differences in parents' attributions for children's school performance in this immigrant multicultural setting, contrary to some previous work with cross-cultural samples (e.g., Hess et al., 1987). However, an analysis of parent decision making style variations (based on the classifications determined above) indicated that authoritative parents were indeed more likely to attribute school performance to the home than were other groups (Pratt & Sebastian, 1989).

In the present analysis we assessed differences among teachers and our three parent decision making style groups in patterns of attribution to home, school, effort, and ability factors. Table 3 shows the means for these analyses.

A Parent Decision Style/Teacher Group (4) MANOVA on the four attribution factors of ability, effort, teacher, and home influences on school performance revealed a significant Group X measures interaction, Pillai-Bartlett Trace $V(9,132)=5.19$, $p<.001$. There was also a significant measures effect, Pillai-Bartlett Trace $V(3,42)=8.71$, $p<.001$, due to lower attribution to the home overall than to all other factors (see Table 3).

Follow-up one-way univariate Group (4) ANOVAs were conducted on each of the attribution factors to examine the interaction above. These revealed no

Table 3
School Performance Attribution Means by Parent-Teacher Groups

	Attribution Factor			
	Ability	*Effort*	*Teacher*	*Home*
Authoritative Parents	1.57	3.64	4.29	2.00
Permissive Parents	3.18	3.55	4.45	0.36
Authoritarian Parents	3.45	3.64	3.18	0.73
Teachers	2.75	2.33	1.17	3.75
Unweighted Means	2.74	3.29	3.27	1.71

significant effects for either ability or for effort attribution means. The group differences for school/teacher attribution were significant, however, $F(3,44) = 10.07$, $p<.001$. Scheffe tests indicated that teachers differed from authoritative and permissive parent styles at the .01 level, but not from authoritarian parents. Teachers viewed their own influence on student performance as lower than did all parent groups (see Table 3). There was also a significant effect for home attributions, $F(3,44) = 8.99$, $p<.001$. Scheffe tests indicated that teachers differed from both permissive and authoritarian groups of parents at the .01 level in seeing the home's influence as more important, but did not differ from authoritative parents.

These results suggest a clear tendency for parents and teachers each to view the other group as more responsible for children's achievement. One interpretation of this finding might be in terms of the phenomenon of actor-observer effects in social psychology (e.g., Fiske & Taylor, 1991). In many situations, the causal role of the other person has been shown to be much more salient than one's own role as an "actor." This difference could also be due to defensively motivated, self-serving biases, such that one avoids taking responsibility for children's problems if these are particularly salient in the attribution context (Fiske & Taylor, 1991). A recent study demonstrated that parents who were less satisfied with children's progress did attribute academic problems less to their own childrearing practices than did those parents who were more satisfied (Himelstein, Graham, & Weiner, 1991).

More investigation of this difference between teachers and parents is needed. Note, however, that teachers were found more closely to resemble parents with a joint, authoritative decision style in their tendency to attribute more responsibility for school performance to the home. In fact, teachers often stressed the essential role of parents in influencing school achievement in many ways. For example:

> I think they [parents] have a lot of influence ... There's the whole level of expectation of the parent. I live in a rural area where parents are really excited when a kid finishes high school and that is the goal. There are very few kids that go on after high school. They've set that level of expectation. And some parents make that mistake—like I'm a ditchdigger so you won't be—but they never go back to school to upgrade so the kid never gets the feeling that they have to be anything better ... And the whole level of attitudinal support is really the parent's responsibility.

Conclusion

The present study had two objectives: to describe junior high school teachers' views on parent decision making styles and parent-school relations, and to see if there was a tendency for these views to be most congruent with those of parents characterized by a more authoritative pattern of parenting adolescents. The results suggest that teachers often differed considerably from all parents in their views about schooling and the home. Nevertheless, they did tend to agree more closely with authoritative parents than with others in several spheres, including ideal family decision making styles, ideals about specific parent homework assistance, and ideas about the causal role of the home in school achievement.

Thus, according to teachers, parents should provide joint decision making opportunities to adolescents, can be most helpful through structuring homework situations and monitoring performance on homework assignments, and should see themselves as having a relatively substantial impact on the child's school achievement. In all these domains teachers in the present study seemed to share with parents with a more authoritative decision making pattern a model of the family and its appropriate modes of coordination with the school. As noted above, greater emphasis on the parent's role in school achievement has been shown to be closely linked to more active homework helping patterns by mothers (Pratt, 1993). However, teachers in the present investigation tended to differ sharply from all parents in terms of opinions about the utility of homework helping in seeing this as a less valuable focus for parent engagement than did parents themselves.

Clearly it is appropriate to be cautious in regard to the generalizability of these findings. The sample of teachers and parents was drawn from a single school system, and sample sizes were quite small. Nevertheless, there was nothing to indicate that this junior high school was in any way atypical of other schools in such urban, multicultural settings. Of course, the investigation needs further replication in other contexts. However, it is notable that teachers' discrepancies with nonauthoritative styles of parenting, and greater agreement with those with authoritative decision making patterns, tended to cut across ethnic variations. This is consistent with the evidence that authoritative parenting is relatively advantaged in relation to children's school achievement across various ethnic groups (Dornbusch et al., 1987), although it is certainly the case that these advantages are stronger for some subcultures than for others (e.g., Darling & Steinberg, 1993).

The finding that teachers tended to be more congruent on several dimensions with those parents with an authoritative decision making style raises an interesting value question. With regard to ethnic diversity, it is of course appropriate to argue for the importance of tolerance and teacher understanding of all parents. However, with respect to variations in parenting patterns, this situation is not so clear. There is considerable evidence, as reviewed above, that authoritative styles are associated with better school achievement and social adjustment in children generally (e.g., Dornbusch et al., 1987; Lamborn, Mounts, Steinberg, & Dornbusch, 1991; Steinberg et al., 1994).

On the other hand, parents who prefer to utilize different socialization approaches surely ought to have considerable latitude to do so, hopefully without their children suffering negative consequences in the school system due to potential teacher bias. Furthermore, there is also evidence that certain parenting styles may be differentially appropriate to different subcultural niches. For example, a more authoritarian pattern may be more appropriate to families rearing children in high-risk social environments (Steinberg, Dornbusch, & Brown, 1992). As noted in the introduction, there is considerable evidence that teacher bias in the area of ethnic or social class background has deleterious consequences for children's school careers (Rist, 1970). Investigation of whether the attitude and value discrepancies between less authoritative parents and teachers found in this research are actually translated into specific

teacher attitudes and behaviors directed toward students or their families (particularly those from different subcultures) is badly needed.

The findings with respect to the homework views of teachers and parents are particularly noteworthy, as homework appears to have considerable positive benefits for student achievement (Keith, Reimers, Fehrmann, Pottebaum, & Aubey, 1986). In early adolescence, homework activities specifically involving the parent seem to be of value (Leone & Richards, 1989). The views of teachers in the present study were decidedly inconsistent with this evidence, however. Although teachers were open to talking with parents about homework issues, they were certainly not as enthusiastic as parents about home participation in these activities. Clearly more consideration needs to be given to this important teacher-parent discrepancy and its possible impact on children's school achievement.

Acknowledgments

The authors thank Ralph Cuthbertson for his help with this research and the parents and teachers who participated for their thoughtful interest. This research was conducted by the second author as part of a master's thesis under the supervision of the first author at Wilfrid Laurier University. Some aspects of the research were supported by a Social Sciences and Humanities Research Council of Canada grant to the first and third authors. Portions of this research were reported at the Society for Research in Child Development Meetings, Kansas City, April 1989.

References

Baumrind, D. (1973). The development of instrumental competence through socialization. In A. Pick (Ed.), *Minnesota symposium on child psychology* (vol. 7). Minneapolis, MN: University of Minnesota Press.

Baumrind, D. (1991). Effective parenting of adolescents. In P. Cowan & E.M. Hetherington (Eds.), *The effects of transitions on families*. Hillsdale, NJ: Erlbaum.

Cazden, C. (1988). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Chavkin, N. (1990). A multicultural perspective on parent involvement: Implications for policy and practice. *Education, 109*, 276-285.

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin, 113*, 487-496.

Dornbusch, S., Carlsmith, J.M., Bushwall, S., Ritter, P., Leiderman, H., Hastorf, A., & Gross, R. (1985). Single parents, extended households, and the control of adolescents. *Child Development, 56*, 326-341.

Dornbusch, S., Ritter, P., Leiderman, P. H., Roberts, D., & Fraleigh, M. (1987). The relation of parenting style to adolescent school performance. *Child Development, 58*, 1244-57.

Fiske, S., & Taylor, S. (1991). *Social cognition* (2nd ed.). Reading, MA: Addison-Wesley.

Heath, S. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge: Cambridge University Press.

Hess, R., Chih-Mei, C., & McDevitt, T. (1987). Cultural variations in family beliefs about children's performance in mathematics. *Journal of Educational Psychology, 79*, 179-188.

Himelstein, S., Graham, S., & Weiner, B. (1991). An attributional analysis of maternal beliefs about the importance of child-rearing practices. *Child Development, 62*, 301-310.

Keith, T., & Page, E. (1985). Homework works at school: National evidence for policy changes. *School Psychology Review, 14*, 351-359.

Keith, T., Reimers, T., Fehrmann, P., Pottebaum, S., & Aubey, L. (1986). Parental involvement, homework, and TV time: Direct and indirect effects on high school achievement. *Journal of Educational Psychology, 78*, 373-380.

Lamborn, S., Mounts, N., Steinberg, N., & Dornbusch, S. (1991). Patterns of competence and adjustment among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development, 62*, 1049-1065.

Leone, C., & Richards, M. (1989). Classwork and homework in early adolescence: The ecology of achievement. *Journal of Youth and Adolescence, 18*, 531-548.

Meadows, S. (1993). *The child as thinker*. London: Routledge.

Paulson, S. (1994). Relations of parenting style and parental involvement with ninth-grade students' achievement. *Journal of Early Adolescence, 14*, 250-267.

Pratt, M. (1993, June). *Parental attributions for children's school achievement and parents' schooling-related practices*. Poster presented at the Canadian Psychological Association Meetings, Montreal.

Pratt, M., Green, D., MacVicar, J., & Bountrogianni, M. (1992). The mathematical parent: Parental scaffolding, parenting style, and learning outcomes in long-division mathematics homework. *Journal of Applied Developmental Psychology, 13*, 17-34.

Pratt, M., Hunsberger, B., Pancer, S. M., Roth, D., & Santolupo, S. (1993). Thinking about parenting: Reasoning about developmental issues across the lifespan. *Developmental Psychology, 29*, 585-595.

Pratt, M., & Sebastian, T. (1989, April). *Parental beliefs and practices regarding school and homework: Parenting style and ethnic differences at the junior high level*. Poster presented at the Society for Research in Child Development Meetings, Kansas City.

Rist, R. (1970). Student social class and teacher expectations: The self-fulfilling prophecy in ghetto education. *Harvard Educational Review, 40*, 411-451.

Steinberg, L., Dornbusch, S., & Brown, B. (1992). Ethnic differences in adolescent achievement: An ecological perspective. *American Psychologist, 47*, 723-729.

Steinberg, L., Elmen, J., & Mounts, N. (1989). Authoritative parenting, psychosocial maturity, and academic success among adolescents. *Child Development, 60*, 1424-1436.

Steinberg, L., Lamborn, S., Darling, N., Mounts, N., & Dornbusch, S. (1994). Over-time changes in adjustment and competence among adolescents from authoritative, authoritarian, indulgent, and neglectful families. *Child Development, 65*, 754-770.

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 188-212*

David R. Jones

and

Edward A. Holdaway
University of Alberta

Expectations Held for Department Heads in Postsecondary Institutions

Current changes in mission, mandate, organizational structure, funding, and student demographics evident in postsecondary institutions reinforce a general recognition of the importance of capable departmental leadership. This study examined managerial, academic leadership, and politician/advocate roles of department heads in a community college, a technical institute, and a university in Alberta. Information regarding commonalities and differences in activities, personal traits, and administrative skills perceived relevant to these roles was obtained from interviews (20) and questionnaires (160 respondents—deans, incumbent department heads, and faculty association executive members). The findings revealed substantial variation among roles for the three types of institutions and the need for administrative skill development appropriate to different stages of service. In addition, the need emerged for a greater understanding of role functions by senior administrators, department heads themselves, and particularly by those who provide preservice and inservice education for department heads.

Les changements récents dans les domaines de la mission, du mandat, de la structure organisationelle, de financement, et de statistiques démographiques des étudiant(e)s sont évidents dans les institutions postsecondaires. Ces changements confirment l'importance déjà connue d'un leadership de qualité au sein d'un département. Dans cette étude on examina les rôles des cadres, du leadership académique, et les rôles politiques des chefs de département dans un collège communautaire, un institut polytechnique, et une université en Alberta. L'information concernant les points communs et les différences d'activités, les traits personnels, et les habilités administratives perçus comme étant pertinents à ces rôles a été obtenue d'entrevues (20) et de questionnaires (160 répondant(e)s—doyens et doyennes, nouveaux et nouvelles chefs de département, et des membres des comités exécutifs des associations de faculté). Les résultats de la recherche indiquent une variation importante parmi les rôles des trois sortes d'institutions et un besoin de développer les habilités administratives appropriées aux différents stages de service. En plus a ressorti le besoin d'une meilleure compréhension des fonctions des rôles des administrateurs supérieurs et des administratices supérieures, des chefs de département eux-mêmes et elles-mêmes, et particulièrement des responsables des services de formation pédagogique pour les chefs de département.

David Jones is a sessional lecturer in the Department of Educational Policy Studies. His recent research activities have involved outreach delivery of undergraduate programs and extension function in universities.

Edward Holdaway is a professor of educational administration. His academic interests include graduate studies, program review, institutional research, job satisfaction, and effectiveness.

Introduction

It is widely recognized that higher education is a complex and unique administrative domain. Clark (1983) asserts that the higher education system is fraught with

> inordinate and uncommon complexity ... [and] to understand that complexity much better than we currently do requires that we retreat somewhat from general theorizing across the major organized sectors of society [or institutions] and concentrate on analysis of particular realms. But to begin from the assumption of other sectors is to misperceive and underestimate the unusual parts in the mixture of the common and the unique. (p. 276)

He adds that "growing complexity increases our uncertainty" and that an understanding of expectations will enhance our potential to find or develop new means of dealing with common problems (p. 270). The academic department is such a complex realm and the department head is an often "misperceived and underestimated" part.

The perception of department heads (department chairs, program coordinators) as generic office holders with responsibilities that vary little, regardless of the institution or discipline, is not uncommon. Such a perception seems incongruous given differences among the missions and mandates of institutions of higher education. This article reports findings from a recent study that explored the extent to which expectations held for departmental leaders varied in three different types of institutions in Alberta. Different sources of expectations were identified and studied in order to identify institutional variations. A comprehensive and generic model of department head activities, traits, skills, and institutional factors was developed that allows comparisons to be made across disciplinary and institutional boundaries. Furthermore, identification of differences in expectations can be of value in the selection, administrative development, and evaluation of administrative performance of postsecondary department heads.

Role Expectations

Located at the point of product delivery, department heads are the "interface" with central administrators on behalf of faculty and with faculty on behalf of administrators. Tucker (1992) states that the generic purpose of department heads is to "provide leadership to the faculty and at the same time to supervise the translation of institutional goals and policies into academic practice" (p. vi). The forward-looking focus demanded by institutional imperatives and the inertia of academe often generates a push-pull dynamic that creates a need for department heads to perform different roles in the service of administration versus faculty interests. These imperatives describe a position fraught with role ambiguity and potential for conflict among the many expectations held for administrative behavior. Eley (1994) considers that the current role of United Kingdom university department heads is substantially different from what it was in the 1960s and 1970s.

Contributing to the complexity of the position is the assertion by some scholars that much of the decision making in postsecondary institutions occurs at the departmental level. In a pioneering but still frequently cited study, Heimler (1967) estimated that 80% of all administrative decisions occur at this

level. Within departments there is a great potential for competing demands. Bragg (1981) contended that department heads have a daunting task of collating formal or overt expectations, transmitted from job descriptions and mission statements, with the informal or covert expectations transmitted through peer attitudes and departmental climate. Tucker (1992) observed that "department heads are drawn from faculty ranks and have had, at best, very little administrative experience" (p. 27). Given the demands and their limited experience, department heads can be viewed as somewhat disadvantaged. A central difficulty, particularly for new department heads, is sorting out what is needed from what is wanted. The UK *Jarratt Report* (Committee of Vice-Chancellors and Principals, 1985) considers that university department heads should be good managers as well as distinguished academics, with priority being given to managerial capability.

Several role typologies and taxonomies of duties and responsibilities of department heads have been promoted, with that compiled by Tucker (1992) being a widely recognized standard (Baker, 1993). Different arrangements of duties by many experts invariably focus on management, academic, and developmental dimensions of the position (Bragg, 1981; Creswell, Wheeler, Seagren, Egly, & Beyer, 1990; Gmelch, 1992; Moses & Roe, 1990; Moxley & Olson, 1990). Of these, management and academic leadership are traditional roles of department heads. The developmental dimension is cited by Bennett (1983) as an inherently political orientation often focused on advocating departmental initiatives and cultivating relationships and networks. In the current climate of fiscal restraint and critical public scrutiny of postsecondary institutions, developing and expanding networks is necessary.

Demands related to the various dimensions, duties, and responsibilities of the position of department head have been scrutinized by scholars to ascertain the specific traits and skills needed for their performance. For example, a study of Australian university chairs conducted by Lonsdale and Bardsley (1984) identified several inservice training needs. In the early 1980s, major contributions were made by Tucker (1983, 1984) and Bennett (1983) to inservice training of department heads under the auspices of the American Council on Education. More recently, Moses and Roe (1990) outlined several areas of skill development for university department heads in Australia that include peer skills, leadership skills, conflict-resolution skills, information processing skills, decision making skills, resource-allocation skills, and entrepreneurial skills. Although worthy of citation, these studies focus on a single institutional type, and as such provide a limited view of postsecondary department heads.

Many of the foundation role studies of department heads used approaches that can best be described as quantitative or positivistic. This particularly applies to those studies that relied solely on numerical responses to questionnaire items. Authors such as Bensimon, Neumann, and Birnbaum (1989) and Roueche, Baker, and Rose (1989) recommended that a more qualitative approach to this type of study is now desirable.

A study of department heads in different postsecondary contexts by examining roles, activities, related traits, and skills using both quantitative and qualitative approaches was warranted. Such a study required a definitional framework that would be applicable across disciplinary and institutional

boundaries. *Role*, although a common term, is variously applied. Heiss (1981) commented that the "disparate, confusing and arbitrary" use of this term is a significant problem associated with the concept of roles (p. 94). However, an operating definition can be credited to Zurcher (1983) who considers roles as behavioral "should-do's" and expectations as institutionalized shared understandings of roles.

Other key terms associated with roles require working definitions. For example, Biddle (1979) defines *position* as "a designation familiar to those in a subject population" (p. 395). Subsumed by this definition are categories of behaviors that have a narrower focus. These behaviors or *functions* were defined by Biddle as "goal-related results of behavior" (p. 389). An interpretation of Biddle's definition of activities are behaviors that may be characteristic of more than one role and function. For example, planning activities are equally applicable to financial management, personnel development, and curriculum design. How these framework terms were applied in this study is manifest in the questionnaire design as exemplified in the format used in Table 1; activities are listed under function categories by role.

Method

The main research question was as follows: To what extent do expectations for department heads differ among types of postsecondary institutions in Alberta? Three different types of institutions agreed to participate in the study: a community college, a technical institute, and a university. These institutions were chosen because they would foster the greatest potential for comparison of behavioral expectations. Furthermore, it was anticipated that the study would show areas of similarity among department heads in the three institutions.

Sample

All three institutions are located in an urban setting. However, each serves different student markets, has a distinct academic mission, and a unique history. The university and the institute were founding pillars of the postsecondary system in Alberta, becoming operational in 1908 and 1916. From 1916 to 1982 the institute was directly administered by the government, whereas the university had a board of governors. In 1970 the community college was established with a board of governors. The institute was realigned in 1982 with a board governance model. Although now all are board-governed, some current differences in governance, administrative structure, and practice among the three institutions are worthy of note.

The university has a board of governors, a general faculties council (which makes academic policy decisions, in some cases subject to board approval), and a senate (which serves as a link between the community and the university). In contrast, the college and the institute have boards of governors and academic councils, but no senates. The academic councils in Alberta's postsecondary institutions vary in the degree of authority they possess with respect to approval of academic matters.

The institute and the college, although similarly structured and in some ways imitative of the university's governance and administrative structure, exhibit differences in practice. Differences in mission, student demographics, and history have shaped many of their institutional norms. This study dealt

with manifestations of these norms at the departmental level, in particular those related to department heads, such as the titles given to first-line administrators. The collegial title *department chair*, common to both the college and university, contrasts with the authoritarian impression generated by *program coordinator* or *director* and the generic label *academic manager* used by the technical institute. Although perhaps a superficial example, titles can generate differences in expectations held for administrative behavior. In fact, several interview comments identified the *academic manager* title and the still prevalent controlling *management* ethos as residues of the period when the institute was governed directly by the provincial government.

Despite these differences, some isomorphism exists among the three institutions. This is attributable to their involvement with the same transfer and outreach programs; similar reliance on provincial operations, capital funding, and student support programs; and educational background and experience of their administrators. In order to acquire an overview of the expectations influenced by these similar and dissimilar aspects, three constituent groups were surveyed in each institution: deans, incumbent department heads, and faculty association executive members. This latter constituent group was chosen because it represented a relatively expert, broadly representative, and institution-wide perspective.

Instruments

Development of the survey questionnaire and interview guide was assisted by a series of preliminary interviews, a review of pertinent literature, and thorough pilot testing of their content and formats. The preliminary interviews were conducted with a vice-president (academic), two deans, and two department heads.

Questionnaires were sent to all deans, incumbent department chairs, and faculty association executive members. They were requested to indicate the importance ("not important," "fairly important," "very important," or "essential") that they placed on various role-specific activities, traits, skills, and institutional factors. Samples of the questionnaire are included in the Appendix. The 160 responses (66.7%) obtained were as follows: community college (35), institute (50), and university (75). Of these responses 20 were from deans, 17 from faculty association executive members, and 123 from incumbent chairs. Interviews were conducted with 17 department chairs from different academic divisions and with one vice-president (academic) or associate vice-president (academic) in each institution.

Analyses of the survey data and content analysis of open-ended responses and interview data were performed and are reported in this article. The comments included have been slightly paraphrased where necessary to improve readability without altering meaning. This article does not address in detail the matter of authority attributed to different roles.

Follow-up Interviews

Follow-up interviews were intended to provide a personalized view of issues and to allow for necessary clarification of responses to the questionnaire responses. They were conducted with a purposive sample of 17 department chairs, two vice-presidents (academic), and one associate vice-president

(academic). All department chairs interviewed were identified by the written authorization indicated on returned reply pages distributed with each questionnaire.

Document Sources

Pertinent information was extracted from documents related to department chairs from the institutions and other sources. These documents were primarily used to provide background information and were not analyzed for content.

Reliability and Validity

Maguire (1990) asserts that "validity of method is not merely a matter of technique properly applied, it is an aid to credibility" (p. 298). In this study efforts were made to promote credibility by clarifying the problem through preliminary interviews, a review of relevant literature, and piloting the design and content of the instruments.

The sample chosen for this study, although largely purposive, nevertheless may justify some claim to transferability to other postsecondary institutions, if only in the same postsecondary system. However, the descriptive character of this study greatly limited any guarantee of replication of the findings. The carefully conceived nature of the research instruments and their rigorous application generated a measure of reliability.

An attempt was made to reduce undue researcher bias and influence during data collection. This was of particular importance because the researcher had previous experience as a department head in the postsecondary domain. With this concern in mind, feedback on the preliminary analysis of questionnaire responses was obtained during interviews. The use of verbatim quotations from interviews and information from institutional documents aided in establishing the validity of the findings.

Findings

Given the purpose of examining the extent of differences in expectations for department head administrative behavior, only substantial differences—15% or greater—between any pair of aggregate or frequency distribution response percentages are reported. This criterion was met by 46 items out of 78 (59%). These aggregate percentages consisted of "very important" and "essential" responses, which are the percentages shown in Tables 1-5.

Roles and Activities

In this section the responses for activities categorized under each of the roles are presented.

Managerial role. Responses generally indicated comparatively higher expectations in financial and control activities for institute department heads. The college respondents, however, placed more importance on operations and communication activities. From the financial management activities shown in Table 1, nearly identical percentages were obtained from the university (76%) and college (77%) for budget development and the same percentage (79%) for budget control. However, the percentages from the institute respondents were 98% and 96% for these two activities. A similar pattern emerged for the percentages obtained for the operations management activity of monitoring, with

Table 1
Aggregate Percentages of "Very Important" and "Essential" Ratings of Selected Department Head Activities, Classified by Role, That Showed Substantial Differentiation

Roles and activities	*College* %	*Institute* %	*University* %
Managerial			
Financial management			
Budget development	77	98	76
Budget control	79	96	79
Operations management			
Monitoring	65	82	66
Scheduling	91	63	51
Information dissemination	94	90	60
Academic leadership			
Program activities			
Curricular (design and delivery)	89	64	80
Personal academic activities			
Teaching	94	58	70
Research	33	28	86
Writing	46	35	76
Consulting	48	50	30
Professional affiliations	82	67	4
Politician/advocate			
Public relations activities			
Business community	59	92	33
Internal committee activities			
Collective bargaining	48	58	19
Salary and promotion	65	61	94
External activities			
Professional associations	90	75	51
Government programs	45	76	40
Corporate initiatives	50	71	36

Notes: 1. The numbers for each institution in each of Tables 1-5 were College (35), Institute (50), and University (75).
2. "Substantial differentiation" means that any one pair of responses differed by 15% or more.
3. See question 10 in Appendix.

the institute (82%) showing marked differentiation from the college (65%) and the university (66%).

Wider differences were recorded for scheduling, with the college (91%) providing the higher percentage compared with the institute (63%) and the university (51%). College (94%) and university (60%) responses showed a similar difference for information dissemination.

One respondent considered that the "dean's directives influence [department heads'] operations as a manager in resource areas." The reference to reduced authority was reinforced by another department head who concluded that "little scope [exists] for discretion within budgets ... largely a pro forma exercise within the unit." However, another emphasized an evolving aspect of the role: "responsibilities of department heads have changed dramatically in the past 5-7 years; financial management and personnel management are now very important." This was reinforced by the statement that "the chair's role is chiefly personnel management."

Academic leadership role. Marked differences were evident in the programming function of this role and in expectations for personal academic performance. Of the program activities curriculum design and delivery provided the largest percentage difference between the college (89%) and the institute (64%).

Also, several striking differences emerged among the personal academic activities of department heads. Teaching received a substantially higher response from the college (94%) than from either the institute (58%) or the university (70%). Likewise, the college registered a higher percentage (82%) for maintaining professional affiliations compared with the institute (67%) or the university (47%). Not surprisingly, percentages obtained from the university for research (86%) and writing (76%) were considerably higher than those obtained from either the institute (28% and 35%) or the college (33% and 46%). By contrast, consulting received a higher response percentage from the institute (50%) and the college (48%) than from the university (30%).

One university department head observed, "it is difficult for any chair to provide academic leadership to all disciplines in a department with varied program content." Another considered that "faculty members need to feel credibility in a chair's teaching ability." With respect to program, one university respondent stated that department heads must "ensure standards, especially in professional disciplines where response to accrediting agencies, field placement, and job market is key."

Politician/advocate role. The percentages obtained for politician/advocate activities showed some of the greatest differences. For example, public relations efforts focused on the business community received a substantially higher percentage from the institute (92%) than from the college (59%) and especially from the university (33%). Department head committee advocacy related to salary and promotion showed a similar, although opposite, percentage difference between the university (94%) compared with the college (65%) and the institute (61%). Collective bargaining advocacy, on the other hand, was a higher expectation for institute department heads (58%) than for those from the university (19%). Several department head advocacy and promotion activities external to their institutions were accorded substantially different percentages. For example, liaison with professional associations was considered very important or essential by more college (90%) than institute (75%) or university (51%) respondents. However, higher percentages were obtained for advocacy efforts focused on government programs and corporate initiatives from the institute (76% and 71%) compared with the college (45% and 50%) or the university (40% and 36%). An interesting comment was, "I suspect the

politician/advocate was a surprise role for many new chairs ... it is neglected at your peril."

Personal Traits and Administrative Skills

Many percentages obtained with respect to the importance of personality traits and administrative skills were either identical or similar for the combined responses of "very important" and "essential." For example, percentages of 89% or greater were obtained from all three institutions for being trustworthy, being collaborative, and being decisive. Similarly, communication, problem solving, and time management received 92% or greater responses as very important or essential administrative skills. However, two traits and one skill that are frequently cited as crucial to administrators in the 1990s—being a risk taker, being entrepreneurial, and being computer literate—garnered strikingly different and generally lower percentages (see Table 2). The combined "very important" and "essential" percentages accorded these traits and skills were lowest for the university sector. Specifically, "being a risk taker" was seen to be "very important" and "essential" by 81% of college respondents but only by 48% of university respondents. The following two comments apply to all department heads regardless of type of institution:

> One must be trusted, respected by colleagues, and action oriented.
>
> Working with various groups, in and out of the department, to solve problems and coordinate action are critical responsibilities, and a sense of humor helps in survival.

The responses about the importance of personal traits and administrative skills were also analyzed by type of role. In general, Table 3 includes much lower percentages of "very important" and "essential" responses regarding nearly all personal traits and administrative skills in relation to the politician/advocate role as compared with the other two roles. Of particular interest are the extremely large differences between the percentages obtained for administrative skills relative to the managerial role compared with the politi-

Table 2
Aggregate Percentages of "Very Important" and "Essential" Ratings of Selected Personal Traits and Administrative Skills of Department Heads That Showed Substantial Differentiation

Personal traits and administrative skills	*College %*	*Institute %*	*University %*
Personal traits			
Being a risk taker	81	68	48
Being entrepreneurial	60	70	51
Administrative skill			
Computer literacy	53	58	39

Notes: 1. "Substantial differentiation" means that any one pair of responses differed by 15% or more.
2. See question 16 in Appendix.

Table 3
Percentage Frequency Distributions of Selected Personal Traits and Administrative Skills of Department Heads Rated as Being Crucial to Different Roles

Personal traits and administrataive skills	*Managerial role*			*Academic leader role*			*Politician/ advocate role*		
	Coll. %	*Inst. %*	*Univ. %*	*Coll. %*	*Inst. %*	*Univ. %*	*Coll. %*	*Inst. %*	*Univ. %*
Personal traits									
Being trustworthy	97	86	89	91	84	84	77	52	56
Being future-oriented	57	66	37	89	68	73	46	50	33
Being collaborative	86	68	69	86	72	69	49	28	28
Being decisive	94	88	77	63	66	47	46	24	24
Having a sense of humor	83	74	72	74	68	56	60	40	39
Being a risk taker	57	58	37	63	38	51	40	42	27
Having collegial respect	74	64	69	74	92	84	54	28	35
Being entrepreneurial	60	64	47	46	38	24	34	34	41
Administrative skills									
Problem-solving	97	88	82	63	58	57	34	22	16
Time-management	86	86	87	57	50	37	17	14	11
Communication	86	82	85	86	80	77	71	44	64
Coordination	77	74	85	60	66	49	26	20	27
Stress-reduction	77	76	68	57	46	36	23	12	5
Computer-literacy	51	60	52	40	50	28	3	6	3
Conflict-management	77	80	83	68	56	55	23	20	19
Small group facilitation	69	68	63	77	64	52	23	22	13
Public relations	60	64	40	43	50	45	66	58	71

Note: See question 17 in Appendix.

cian/advocate role, in some cases over 60%. For example, the percentage differences obtained from all three institutions for application of problem solving skills being either "very important" or "essential" in the managerial role compared with in the politician/advocate role were from the college (97% vs. 34%), the institute (88% vs. 22%), and the university (82% vs. 16%). Two notable exceptions were communication and public relations, where the differences were either not as great or were reversed. The percentages obtained for "very important" or "essential" personal traits for the politician/advocate role were also generally lower than for the other two roles, but the differences were less marked than for administrative skills.

Institutional Factors

Department head job descriptions, performance reviews, and statements of goals and objectives produce overt expectations for performance. These are augmented by less explicit or covert expectations generated by institutional factors such as selection criteria, titles, terms of service, and the disciplinary nature of departments, which are of particular research interest because of their relative transparency.

The highest percentages as sources of expectations for department heads were obtained for overt expectation factors: departmental objectives (college 94%, institute 84%, and university 93%) and regular performance reviews (college 77%, institute 71%, and university 75%). However, as shown in Table 4, a number of factors associated with covert expectations showed greater differentiation and consistently lower percentages. These covert factors were related to selection/election and renewal of department head in the college, the title and discipline alignment in the institute, and the duration of term of service as department head.

Percentages obtained from college respondents regarding election of department heads by peers (56%) and renewable terms of office (64%) were substantially higher than the percentages obtained from the institute (35% and 40%) or the university (42% and 44%) for these factors. A higher percentage of institute respondents rated title of position (46%) higher in importance than did those from the college (31%) and the university (21%) sectors. However, the institute (63%) and the college (61%) had nearly identical percentages of importance for departmental discipline compared with university (43%). Duration of term of service was rated as a key source of expectations for department heads by considerably more respondents from both the university (53%) and the college (50%) than from the institute (33%).

Selection and Review Criteria

Considerable similarity in percentages for "very important" and "essential" responses was obtained regarding teaching experience as a criterion for selecting department heads—college (82%), institute (71%), and university (74%)—and administrative experience as a criterion for their review—college (94%), institute (88%), and university (90%). As shown in Table 5, considerable differentiation was found between the university (75%), the college (14%), and the institute (23%) regarding the aggregate "very important" and "essential" responses for research interests. However, the college (66%) contrasted with the institute (44%) and the university (39%) concerning academic specialty as a selection criterion. A parallel differentiation in percentages was obtained for

Table 4
Aggregate Percentages of "Very Important" and "Essential" Ratings of Selected Institutional Factors That Showed Substantial Differentiation as Sources of Expectations for Department Heads

Institutional factors	*College* %	*Institute* %	*University* %
Election of department heads by peers	56	35	42
Renewable terms of office	64	40	44
Title of position	31	46	21
Department discipline	61	63	43
Duration of term of service	50	33	53

Notes: 1. "Substantial differentiation" means that any one pair of responses differed by 15% or more.
2. See question 18 in Appendix.

Table 5
Aggregate Percentages of "Very Important" and "Essential" Ratings of Selected Selection and Review Criteria That Showed Substantial Differentiation as Sources of Expectations for Department Heads

Selection and review criteria	*College %*	*Institute %*	*University %*
Selection criteria			
Research interests	14	23	74
Academic speciality	66	44	39
Review criteria			
Scholarly performance	42	42	75
Teaching performance	79	49	56

Notes: 1. "Substantial differentiation" means that any one pair of responses differed by 15% or more.
2. See question 18 in Appendix.

scholarly performance from the university (75%) compared with the college and the institute (both 42%), and from the college (79%) versus the university (56%) and the institute (49%) regarding teaching performance as review criteria. Several respondents commented on the need to use different performance criteria for different circumstances. One was critical in noting, "acceptable performance was related to the dean's style of administration."

Summary and Discussion

Traditionally, the indispensable activities of department heads have been associated with the roles of manager and academic leader. Most acknowledged experts have heavily weighted their attention toward these roles (Bennett, 1983; Bennett & Figuli, 1990; Creswell et al., 1990; Emmet & Bennett, 1986; Heimler, 1967; Moses & Roe, 1990; Tucker, 1992). Few have paid attention to differences in the position of department heads in various types of postsecondary institutions. That is, Clark's (1983) assertion quoted at the beginning of this article about the need to "concentrate on analysis of particular realms" has been largely ignored to date with respect to department heads. However, interview findings highlighted the evolving dimensions of the position of department head. Many of the cited changes were connected to a need to add a focus beyond departmental or even institutional boundaries.

Similarities and Differences

Department heads in each of the three different types of postsecondary institutions tended to rate the role activities listed below as either "very important" or "essential" to a greater extent than did department heads in the other two types: college—scheduling, teaching; institute—budget development, budget control, monitoring, public relations with the business community, and activities involving government programs and corporations; university—research, writing, and salary/promotions committees. In general, these differences show greater importance for these department head activities:

teaching at the college; financial control and external relations at the institute; and research and scholarship at the university.

With respect to personal traits and administrative skills, the university respondents tended to rate being a risk taker, being entrepreneurial, and being computer literate substantially lower in importance than did the other two groups. University respondents also tended to rate all the listed administrative skills, except conflict management, lower in cruciality for at least one of the managerial, academic leader, or politician/advocate roles. Similar findings occurred for personal traits, except that for the academic leader and politician/advocate roles the lower ratings were commonly shared by the institute and university respondents.

Regarding the importance of institutional factors influencing department heads, the university and institute respondents tended to have lower percentages of importance ratings, except for duration of term of service (university) and title of position and department discipline (institute). For the importance of selection criteria, university respondents tended to rate research interests higher than did the other two groups, whereas college respondents favored academic specialty. Concerning review criteria, university respondents more frequently selected scholarly performance and college respondents teaching performance.

Some of these findings were predictable, for example, the university group's emphasis on research and writing and the institute group's, given its history, on more managerial control. Other findings were less predictable and warrant further exploration, for example, the college group's emphasis on collaboration as a personal trait that it rated more highly on all three roles.

Tucker (1992) asserted that "departments do not exist in isolation" and that external interaction often "comes to the department whether it seeks it or not" (p. 494). Political aspects related to such interaction and increased visibility can vary by department and institution. One university respondent pointed out the need to be "sensitive to the political dimension of conflict of interest, quality control, and practical ethics issues." Although the importance of these issues on campus can differ by department size and discipline, the off-campus publicity given such concerns can be damaging to reputations and programs. In contrast, an internal political orientation was a typical focus of college respondents, one of whom cited a need for "vigilant awareness of inter-departmental resource allocation equity." Political activity was perceived by several institute respondents as typical organizational bickering. However, comments from institute respondents cited an external focus as being fundamental to any department head's success in generating teaching contracts and developing partnerships with industry. According to several respondents, the profile of the institute in various industry sectors is largely based on the department heads' efforts.

Fundraising campaigns, liaison with accrediting agencies, relations with other institutions and government, as well as cooperative initiatives with business, the recruiting of staff and students, and the placement of graduates were identified by questionnaire respondents as activities that required department heads to have a network of contacts external to their institution. Some of these areas were cited as traditional responsibilities of central administrators, but to

be successful increasingly required the initiative and ongoing involvement of department heads. Public relations, marketing skills, intraorganizational "street smarts," and an extraorganizational vision of trends and issues that could potentially influence the well-being of a department were capabilities perceived by a substantial percentage of respondents as increasingly necessary for department heads. Many of these activities required skills common to politicians and lobbyists. In the opinion of one respondent, "given our current funding situation, especially if a program is under threat, it is a contradiction to undervalue politician/advocate activities."

With an expansion of traditional expectations to include those common to such a politician/advocate role, the position of department head has an added nontraditional dimension. Nonetheless, the findings pointed to a lack of understanding of how a politician/advocate role fits the traditional conceptualization of department head. A generally disparaging aura surrounded references to politics, which obscured for many the benefits of planned and promotive involvement by department heads in public relations, internal positioning, external marketing, and general departmental advocacy activities.

Evolution

Deans were cited in the interviews as perceiving interaction with other institutions and the business community as primary targets for public relations efforts, whereas the campus community was the preferred focus of faculty members and incumbent department heads. Institutional responses exposed similar differences. Public relations activities with the business community and other institutions were considered a higher priority by college and institute than by university respondents.

Internal lobbying for operating and capital resources was viewed by college and institute interviewees as a key activity. Involvement by department heads in salary and promotions deliberations was similarly perceived at the university. Advocacy efforts on behalf of a department or discipline and external lobbying in general were valued more by deans and department head respondents than by faculty members. Similarly, the percentages obtained from college and institute respondents indicated that they were more externally focused than were the university respondents.

As the postsecondary domain changes to include more women and minorities, as well as mature and diversely motivated students, department heads will need to be more outwardly focused and increasingly forewarned of issues and concerns that either pose a threat to or can benefit department disciplines and mission. The reeducation of faculty members regarding these new priorities may be a natural by-product of such an outward advocacy focus. The existence of a capable political/advocacy role model for department heads and, more importantly, development of skills and abilities inherent in such a role will probably be of increasing importance to departments.

Among current dynamics affecting departments are the limitations imposed by constrained resources. An example is the impact on department heads from the merging of departments with diverse mandates. Such mixtures of specialties require a department head to become more of a generalist than would otherwise be necessary in a department with a homogeneous disciplinary mandate. In addition, social and legal issues such as faculty association

bargaining, gender and racial equity, negligence and liability, and political correctness were cited in interviews as taking time and attention away from traditional academic dimensions of departmental leadership.

Such increasing complexity of the position of department head was a concern stated in several interviews. The evolving demands and related expectations for performance warrant a reevaluation of the status and authority of the position of department head in postsecondary institutions. Moreover, with mergers of departments, reduced resources, downsizing of administrative and support staff, and the growing need to be oriented beyond the perimeters of their unit, a reevaluation of appropriate personal characteristics and skills needed for effective performance in the position seems to be in order.

Professionalization

With the increasing complexity in the administration of higher education has come a growing dissatisfaction with the "temporary and amateur" image long associated with departmental leadership. In a recent address to an international conference, Gmelch (1992) concluded that "the time of 'amateur administration' where professors temporarily serve as department chairs, is over" (p. 15). Although we concur in his concern regarding the perception of department chairs as amateurs, Gmelch's conclusion is difficult to support. Given the widespread use of term appointments in North America, we contend that any substantial reinvention of the position will not be easily or quickly accomplished.

Traditional images of department heads and interpretations of appropriate behavior were pervasive in the findings. For example, trustworthiness, collaborative behavior, and collegial respect were identified by high percentages of respondents either as "very important" or "essential" department head traits. However, department head traits related to risk taking or entrepreneurism received substantially less support, especially from the university respondents. A higher percentage of the college and institute respondents valued being future-oriented and decisive than did those from the university.

Substantial differences were apparent regarding most traits as they applied to different roles. For example, even trustworthiness was considered of substantially less value to the politician/advocate role than to a managerial or academic leadership role. Such responses relate to the above-mentioned lower valuation of nontraditional aspects of the department head position. A similar pattern of responses was obtained with respect to administrative skills.

Responses indicated that several people-related skills were highly valued, whereas nontraditional skills were substantially less valued. Of these, communication, problem solving, and conflict management were highly valued, whereas computer skills and wellness-related or stress-reduction abilities were perceived to be of lesser utility. These latter skills are of note because they are those for which specific inservice training is often requested and therapeutically required for administrators in other domains. Furthermore, these and other skills may be some of those on which administrative development programs for postsecondary administrators need to be focused.

The impact on department heads of institutional factors such as position title, type and length of appointments, departmental design, scope of author-

ity, and utilization of department heads in institutional policy formulation have received some attention in the literature. With the exception of the latter item, these factors were given little recognition by respondents in this study as having a noticeable influence on expectations held for department heads.

A critical assessment of department head selection and administrative performance procedures and criteria is needed to ensure that a recognized standard for the position has institution-wide acceptance. A preference for relating job descriptions to evaluations of department head administrative performance was not overwhelmingly apparent in the findings, but it was identified in several interviews as a logical linkage. The endorsement of such a linkage by college department heads was a notable exception to this finding.

Several open-ended questionnaire comments indicated that resistance to a connection between these criteria was based on the reactive nature of the position that frequently defies prescriptive taxonomies of duties and actions. However, a generic conceptualization or position description is a possible alternative. Based on such a generic description, a short-term performance contract or periodic description of objectives, focused on mutually agreed-on projects or objectives, could provide a flexible standard for the evaluation of administrative performance in different contexts. This would benefit deans, who could become acquainted to a much greater degree than was indicated in the interview findings with the dimensions and demands of the department head position. This more formalized system could provide a less isolated working environment for department heads and make the process of performance review less dependent on the personality or stylistic preferences of deans, as alluded to in questionnaire comments and interviews.

To be appointed because of teaching and research accomplishments, but to be evaluated on administrative performance for which one has had little if any preparation is not an arrangement that fosters security. Teaching experience was consistently recognized by a high percentage of respondents as a key selection criterion, as was administrative performance as a review criterion. The preference of university staff for research interests as a selection criterion appeared to be related to their support for scholarly performance as a review criterion.

Performance Review

Several department heads cited the lack of formative administrative evaluation as emanating from undefined criteria. If "professionalism" in this evolving position is required, as Gmelch (1992) asserted, then a concerted critical effort is needed to devise and implement selection and review criteria that are workable across an institution, yet have relevance in each departmental context. Few incumbents have all the requisite skills and abilities fully developed at the time of appointment. Some form of needs assessment is required, as are resources and skill development programming to address the identified needs.

Administrative Development

The recognition of administrative capacities and their development in incumbent and candidate department heads is a concern given little continuing attention by Alberta's postsecondary institutions and virtually no consideration at the higher education system level. Two approaches are available, and

both should be considered: continuing inservice programs and preservice preparation for department head candidates.

Inservice. On-the-job training has been the predominant approach to the administrative development of incumbent department heads. In Alberta's postsecondary system such training has been largely left to individual institutions. However, a system-funded interinstitutional program was offered to university chairs in the maritime provinces in the early 1980s (Baker, 1993), and in recent years multi-day seminars from the Center for Leadership Development—American Council of Education have been offered through Mount Royal College's Extension Department. These latter opportunities have attracted incumbent college and institute department heads, but few if any university chairs. Of additional concern was the lack of attention to uniquely Canadian aspects of higher education administration.

The findings from this study pointed to three categories of inservice skill development needs: (a) institutional administrative procedures, budgeting, and financial skills for "rookie" department heads; (b) individual and small-group mediation skills for second- or third-year department heads; and (c) consensus building, visioning, and planning capabilities as continuing needs. It is obvious that the skills cited are predominantly oriented to management.

With the arguable need for a new focus on external politician/advocate activities comes a concomitant need for the development of skills and capacities related to such a role. Bolman and Deal (1992) asserted that "preservice and inservice programs for school administrators rarely give much attention to symbolic and political skills, yet our results show that they are crucial components for effective leadership" (p. 328). We propose that politician/advocate skills appropriate to each of the categories listed above warrant inclusion in any orientation and inservice administrative development program in higher education.

Preservice. Preservice preparation has been recognized as beneficial to new department heads in making an easy and functional transition from faculty to administrator status. The frequent turnover of department heads resulting from the short-term nature of appointments has reinforced this need. Traditionally, such preparation has taken the form of service as an assistant chair or project coordinator in an academic department. However, one interviewee commented that a formally structured apprentice system ought to be considered, although this form of preservice preparation seems less viable in small or one-person departments that are dependent on part-time staff. A recognized program of preservice courses available to faculty members could be offered on a regular basis through extension services. Participation in a program of this sort could provide a means for senior administrators to identify talent. Program completion could be considered as a selection criterion that provides minimal preparation and gives a better chance for a less stressful transition from faculty to administration for a successful candidate.

Concluding Reflections

This study identified an ongoing set of overt expectations for department heads to be managers and academic leaders in ways traditional to postsecondary education. Traditional notions of department heads as temporary custodians of departmental bastions are being challenged by expectations from

nontraditional students, faculty with nontraditional needs both as scholars and teachers, and by deans who are struggling with the nontraditional demands of their position. Changes in available financial and human resources, disciplinary alignments through mergers and restructuring, demands for alternative delivery, rapid growth in dependence on communication technologies, and a virtual elimination of one-size-fits-all postsecondary education bring increased instability to an environment that thrives on stability. The possibility of department heads continuing to be effective in such a set of circumstances, especially utilizing traditional models and methods, is highly doubtful. A similar view was expressed by Eley (1994) for university department heads in the UK.

Such flux requires department heads increasingly to be specialists, not just in an academic discipline, but in administrative practice. Research and reflection on the covert and overt expectations held for the position of department head can aid departments and deans to react to, and even anticipate, departmental leadership needs. Attention to the expectations inherent in the selection, preparation, and development of motivated and capable candidates for this crucial administrative position in postsecondary education is warranted now more than ever.

References

Baker, R.J. (1993). Chairing the academic department: Leadership among peers. *Canadian Association of University Teachers Bulletin, 40*(1), 12.

Bennett, J.B. (1983). *Managing the academic department.* New York: Macmillan.

Bennett, J.B., & Figuli, D.J. (1990). *Enhancing departmental leadership: The roles of the chairperson.* New York: Macmillan.

Bensimon, E.M., Neumann, A., & Birnbaum, R. (1989). *Making sense of administrative leadership: The "L" word in higher education. ASHE-ERIC Report No.1.* Washington, DC: School of Education and Human Development, George Washington University.

Biddle, B.J. (1979). *Role theory: Expectations, identities and behaviors.* New York: Academic Press.

Bolman, L.G., & Deal, T.E. (1992). Leading and managing: Effects of context, culture and gender. *Educational Administration Quarterly, 28,* 314-329.

Bragg, A.K. (1981). *The socialization of academic department heads: Past patterns and future possibilities.* Paper presented at the annual meeting of the Association for the Study of Higher Education, Washington, DC.

Clark, B.R. (1983). *The higher education system: Academic organization in cross-national perspective.* Berkeley, CA: University of California Press.

Committee of Vice-Chancellors and Principals. (1985). *Report of the steering committee for efficiency studies in universities.* London, UK: Author.

Creswell, J.W., Wheeler, D.T., Seagren, A.T., Egly, N.J., & Beyer, K.D. (1990). *The academic chairperson's handbook.* Lincoln, NE: University of Nebraska Press.

Eley, A. (1994). Management training for the university head of department. *International Journal of Educational Management, 8*(2), 20-22.

Emmet, T.A., & Bennett, J.B. (1986). *Workshop handbook for department and division chairpersons and deans.* Washington, DC: Center for Leadership Development-American Council on Education.

Gmelch, W.H. (1992). UCEA Center for the Study of the Department Chair: Program center update. *University Council for Educational Administration Review, 33*(2), 15.

Heimler, C.H. (1967). The college department chairman. *Educational Record, 48,* 158-163.

Heiss, J. (1981). Social roles. In M. Rosenberg & R.H. Turner (Eds.), *Social psychology* (pp. 94-129). New York: Basic Books.

Lonsdale, A.J., & Bardsley, W.N. (1984). Heads of academic departments: Responsibilities and professional development needs. *Journal of Tertiary Educational Administration, 6*(2), 1-7.

Maguire, T.O. (1990). On methodological sects in educational research. *Alberta Journal of Educational Research, 36,* 297-298.

Moses, I., & Roe, E. (1990). *Heads and chairs: Managing academic departments.* Brisbane: University of Queensland Press.

Moxley, J.M., & Olson, G.A. (1990). The English chair: Scholar or bureaucrat. *Thought and Action, 6*(1), 51-58.

Roueche, J.E., Baker, G.A., & Rose, R.R. (1989). *Shared vision: Transformational leadership in American community colleges*. Washington, DC: Community College Press and American Association of Community Colleges.

Tucker, A. (1983). *Chairing the academic department: Leadership among peers*. New York: Macmillan.

Tucker, A. (1984). *Chairing the academic department: Leadership among peers* (2nd ed.). New York: Macmillan.

Tucker, A. (1992). *Chairing the academic department: Leadership among peers* (3rd ed.). New York: Macmillan.

Zurcher, L.A. (1983). *Social roles: Conformity, conflict and creativity*. Beverly Hills, CA: Sage.

Appendix
Questionnaire Sections

Expectations of Department Heads in Postsecondary Institutions

1. Please check (√) your type of institution.

 1. ☐ community college

 2. ☐ institute of technology

 3. ☐ university

2. What position do you hold at your institution?

3. For how many years have you served in your current position?
 (Count the current year as a full year.) _____ year(s)

4. If you are a department head, how did you receive your appointment? (e.g., election, administrative appointment)

5. What disciplines or program areas are represented by your academic unit? (e.g., faculty, department)

6. What is the predominant type of student registered in courses offered by your academic unit?
 (e.g., undergraduate, graduate, diploma)

7. What percentage of course registrations in your academic unit are in the following categories?

 Diploma _____ %

 Undergraduate _____ %

 Graduate _____ %

8. How many faculty members are employed in your academic unit?

 (a) Full-time equivalent _____ FTE

 (b) Full-time _____ FTE

 (c) Part-time _____ FTE

9. How many associate/assistant department heads are employed in your academic unit? _____

2 — Survey Questionnaire of Expectations of Department Heads in Postsecondary Institutions

10. What level of importance do you place on the following ***managerial*** activities of department heads in your institution?

11. In your institution, what type of authority do you consider that department heads ***have*** and ***should have*** to make final decisions in the following areas?

Please place checks (√) in the appropriate frames opposite each item listed below.

Scale: Not Important — Fairly Important — Very Important — Essential — Not Applicable

Scale: No Authority — Shared Authority — Complete Authority

Not Impt.	Fairly Impt.	Very Impt.	Esstl.	N/A	Managerial Activities	Have			Should Have		
						No Authority	Shared Authority	Complete Authority	No Authority	Shared Authority	Complete Authority
					Financial Management						
					Budget development						
					Budget control						
					Personnel Management						
					Planning						
					Hiring						
					Evaluation						
					Termination						
					Personnel development						
					Facilities, Equipment and Operations						
					Planning (goals, objectives)						
					Resource allocation						
					Monitoring						
					Scheduling						
					Information dissemination						

Comments: _____

12. What level of ***importance*** would you place on the following academic leadership activities expected of department heads in your institution?

13. In your institution, what type of authority do you consider that department heads ***have*** and ***should have*** to make final decisions in the following areas?

Please place checks (√) in the appropriate frames opposite each item listed below.

Scale Not Important — Fairly Important — Very Important — Essential — Not Applicable

Scale No Authority — Shared Authority — Complete Authority

Not Impt.	Fairly Impt.	Very Impt.	Esstl.	N/A	Academic Leadership	Have			Should Have		
						No Authority	Shared Authority	Complete Authority	No Authority	Shared Authority	Complete Authority
					Program Activities						
					Curricular (design & delivery)						
					Program evaluation						
					Academic policy formulation						
					Academic policy implementation						
					Student services						
					Faculty-related Activities						
					Recruitment/Selection						
					Workloads						
					Faculty development						
					Performance review						
					Department Head Academic Activities						
					Teaching						
					Research						
					Writing						
					Consulting						
					Professional affiliations						

Comments: _____

14. What level of importance would you place on the following ***political and advocacy*** activities of department heads in your institution?

15. In your institution, what type of authority do you consider that department heads ***have*** and ***should have*** to make final decisions in the following areas?

Please place checks (√) in the appropriate frames opposite each item listed below.

Scale Not Important — Fairly Important — Very Important — Essential — Not Applicable

Scale No Authority — Shared Authority — Complete Authority

Not Impt.	Fairly Impt.	Very Impt.	Esstl.	N/A	Political and Advocacy Activities	Have			Should Have		
						No Authority	Shared Authority	Complete Authority	No Authority	Shared Authority	Complete Authority
					Public Relations Activities						
					Campus community						
					Inter-institutional						
					Government agencies						
					Elected representatives						
					Business community						
					Committee Activities						
					Capital resources allocation						
					Institutional policy formulation						
					Collective bargaining						
					Operational resource allocation						
					Salary and promotion						
					Entrepreneurial Activities						
					Professional associations						
					Other educational domains						
					Government programs						
					Corporate initiatives						

Comments: _____

16. What level of importance would you place on the following ***personal traits and administrative skills*** of department heads?

17. In which of the three department head roles can the following personal traits and administrative skills be considered ***crucial***?

Please place checks (√) in the appropriate frames opposite each item listed below.

Scale Not Important — Fairly Important — Very Important — Essential — Not Applicable

More than one frame may be checked below

Not Impt.	Fairly Impt.	Very Impt.	Esstl.	N/A	Traits and Skills	Managerial Role	Academic Leader Role	Politician/ Advocate Role
					Personal Traits			
					Be trustworthy			
					Be future-oriented			
					Be collaborative			
					Be decisive			
					Have a sense of humour			
					Be a risk-taker			
					Have collegial respect			
					Be entrepreneurial			
					Administrative Skills			
					Problem solving			
					Time management			
					Communication			
					Coordination			
					Stress reduction			
					Computer literacy			
					Conflict management			
					Small-group facilitation			
					Public relations			

Comments: _____

6 — Survey Questionnaire of Expectations of Department Heads in Postsecondary Institutions

18. What level of importance would you place on the following factors and criteria as ***sources of expectations*** for department heads in your institution?

Please place checks (√) in the appropriate frame opposite each item listed below.

Scale Not Important — Fairly Important — Very Important — Essential — Not Applicable

Factors and Criteria	Not Impt.	Fairly Impt.	Very Impt.	Esstl.	N/A
Institutional Factors					
Institutional mission and goals					
Departmental objectives					
Individualized job descriptions					
Recruiting process of department heads					
Appointment of department heads by superordinates					
Election of department heads by peers					
Duration of term of service					
Renewable terms of office					
Title of position					
Size of department					
Department discipline					
Type of institution					
Regular performance reviews of department heads					
Selection Criteria					
Administrative experience					
Teaching experience					
Research interests					
Academic speciality					
Performance Review Criteria					
Administrative performance					
Scholarly performance					
Teaching performance					
Service activities					

Comments: _____

Thank-you for participating in this survey.

Survey Questionnaire of Expectations of Department Heads in Postsecondary Institutions — **7**

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 213-230*

Myrna L. Greene

and

Carillon R. Purvis
University of Lethbridge

The Forgotten Link: Transitions from Graduate School to Classroom Teaching

Considerable literature addresses the difficult transition from preservice teacher education to the realities of classroom teaching, but there has been almost no research that examines specifically the experiences of experienced teachers returning to the classroom after completing graduate studies. The purpose of this study was to explore the nature of that transition. Using questionnaires (N=50) and in-depth interviews with 17 graduates, we reach a number of conclusions, among them that: graduate studies have a positive impact on the way teachers think about their teaching; teachers who return to the classroom believe that their colleagues resent their "new knowledge"; part-time study appears to lend itself to a more direct questioning of specific practice and an attempt to try out new ideas in the classroom; full-time study results in considerable dissatisfaction with, and almost an inability to cope with, existing systems. A number of recommendations are made for graduate programs and for schools.

Même s'il existe une importante littérature sur les difficultés de transition que ressentent les enseignants et les enseignantes de leurs stages de formation aux réalités de l'enseignement en salle de classe, il ne semble y avoir presqu'aucune recherche qui examine spécifiquement les expériences des enseignants et des enseignantes ayant de l'expérience qui retournent en salle de classe après avoir complété leurs études supérieures. Le but de cette étude était d'explorer la nature de cette transition et de leur retour en salle de classe. En utilisant des questionnaires (N=50) et des entrevues détaillées de 17 gradués universitaires nous avons pu en tirer une série de conclusions, parmi lesquelles: Les études graduées affectent de façon positive la façon que les enseignants et les enseignantes perçoivent leur enseignement; les enseignants et les enseignantes qui retournent en salle de classe croient que leurs collègues éprouvent de l'amertume à l'égard de leurs nouvelles connaissances; l'étude à temps partiel paraît se porter mieux à un questionnement direct de leur pratique spécifique et d'un effort d'essayer de nouvelles idées basées sur la théorie dans leur salle de classe; l'étude à plein temps semble produire une grande insatisfaction des systèmes qui existent et d'une presqu'inabilité de fonctionner avec ces mêmes systèmes. On suggère plusieurs recommandations pour les programmes gradués postsecondaires et pour les écoles.

Most teachers remember their first few months and years of teaching as a stressful, painful blur, and much has been written about the difficult transition from undergraduate student to the first year of teaching as idealism succumbs to the realities of school culture (Greene & Campbell, 1993; Sarason, 1993). There is also research examining changes experienced by teachers who are new to the system (Odell, 1986), and even some that describes change as a function of teacher inservice (McNergney & Carrier, 1981). However, we have been able

Myrna Greene is a professor and Dean of the Faculty of Education.
Carillon Purvis is Assistant Coordinator of Graduate Studies in the Faculty of Education.

to identify little research that examines specifically the experience of moving from teaching into graduate studies and back to classroom teaching. Hargreaves and Fullan (1992) cite research to indicate that simultaneously being a graduate student and teacher "restores meaning and relevance to the graduate programme as a place of productive professional development. Each has its place: the classroom as the busy hub of practice, and the seminar room as a safe refuge for reflection on it" (p. 11). But what happens to the teacher in that "busy hub of practice" when there is no longer time, or a safe place for reflection, or a culture that supports, or even allows, the kind of reflection to which they have become accustomed in their graduate studies?

The MEd program at the University of Lethbridge has as its central focus the professional development of practicing classroom teachers. However, informal comments from some of the graduates who have returned to classroom teaching on completion of their graduate program hint at considerable frustration and difficulty in returning to the classroom, in reconciling the opportunity to think, to reflect, to discuss, and to read with the relentless realities of day-to-day life in schools. According to Duffy and Roehler (1986):

> Teachers restructure new information in terms of their conceptual understandings of curricular content, their concept of instruction, their perception of the demands of the working environment, and their desire to achieve a smoothly flowing school day. As the information is processed through these filters the teachers' thinking changes relative to the innovation. Hence an innovation that is sensible when discussed in a course or an inservice session is modified by the filters; sometimes, the modification is so great that what seemed sensible in the teacher education situation cannot be implemented on a regular basis in the classroom. (p. 57)

Baskett, Marsick, and Cervero (1992) indicate that there are a number of polarities or tensions when professionals continue their education: individual versus collective; rational versus intuitive; cognitive versus emotional; routine versus nonroutine; formal versus informal; constructive versus scientific knowledge (pp. 110-111). We suggest that, although these are not mutually exclusive perspectives, more often than not educators assume models of professional learning that tend to be closer to one side of these constructs than the other. If graduate programs and schools have different assumptions about the purpose of teachers' professional development, then teachers who are returning to the classroom are likely to experience considerable frustration, dissatisfaction, and tension, and the benefits that were expected to accrue as a result of their further education may not be forthcoming. For example, if teachers in a graduate program come to believe that professional knowledge is socially and culturally constructed and content specific, and subsequently attempt to translate those insights into a traditionally structured public school model of learning, they are likely to experience gut-wrenching episodes of disequilibrium. This possibility and the casual comments from former graduate students served as the impetus for this study.

Purpose of the Study

The purpose of the study was to begin to explore the transition from graduate student back to teaching, specifically (a) to describe the impact or effect of

graduate studies on participants as teachers (e.g., their classroom practices, relationships with colleagues, beliefs about teaching, and beliefs and practices about students and student learning); (b) to examine whether the transition experiences were different for different groups of teachers (e.g., full-time/part-time status, gender, age, years of experience; and (c) for teachers who did not return to the classroom, to determine how and to what extent their graduate studies contributed to that career change.

Methodology

This study was conducted in the context of one institution and relied on questionnaire and interview data. Brief questionnaires were developed to address the demographic issues raised in research question (b); open-ended questions addressed research questions (a) and (c). In the open-ended questions respondents were asked to describe their major reasons for applying for graduate studies and the relationship between their graduate studies and their teaching (e.g., beliefs about teaching, teaching practices, staff/student relationships). At the end of the questionnaire respondents were asked to provide a telephone number if they would be willing to participate in a follow-up interview.

Questionnaire Sample

Questionnaires were mailed to all persons who had been in classroom teaching positions on admission to the graduate program and who had graduated between 1984 and 1992 (N=64). Eight questionnaires were returned with the address unknown. From the resulting sample of 56 graduates, 50 (89.2%) of the questionnaires were completed and returned. Of those 50, 44 (88%) agreed to be interviewed.

Demographic information about the study sample is presented in Table 1. Of the 50 respondents, 29 had completed the program entirely part time (10 male, 19 female); 21 had completed at least one semester of the program full time (10 male, 11 female). Although the sample was intended to have been drawn from those who were teaching at the time of admission to the program, it was apparent from their responses that five persons were in "other positions," for example, counselor or resource room teacher, and one was in a purely administrative position. On admission to the program 34 (68.7%) of the participants were in full-time teaching positions. After graduation 22 (44%) were in full-time teaching positions. It is also apparent from Table 1 that most of the respondents were between 30 and 39 years of age, except for the women who had attended part time, of whom 12 of 19 (63%) were more than 40 years of age. The average years of experience ranged from 8.5 years for part-time males to 11.4 years for full-time males.

Interviews

Interviews were conducted with 17 of the 44 graduates who indicated a willingness to be interviewed, in order to explore in greater detail the experience of the transition from graduate study to full-time teaching or to their present position. The initial plan had called for 20 interviews, 10 male and 10 female, with approximately five of each part-time and five full-time. We had also intended that approximately one third would be from the Lethbridge area, one third would be within 160 km of Lethbridge, and one third would live more

Table 1
Demographic Information About Survey Sample

Position	*Part Time (29)*								*Full Time (21)*								*Total*
	Male (10)				*Female (19)*				*Male (10)*				*Female (11)*				*50*
	Tchg	*Tchg/ Admin*	*Admin*	*Other*	*Tchg*	*Tchg/ Admin*	*Admin*	*Other*	*Tchg*	*Tchg/ Admin*	*Admin*	*Other*	*Tchg*	*Tchg/ Admin*	*Admin*	*Other*	
On Admission	8	2			13	3	1	2	7	2		1	6	3		2	50
After Grad	6	1	3		8	6	1	4	3	6		1	5	5		1	50
Age at Admission		<25 - 0 26-29 - 2 30-39 - 7 >40 - 0				<25 - 1 26-29 - 1 30-39 - 5 >40 - 12				<25 -0 26-29 - 0 30-39 - 8 >40 - 2				<25 - 1 26-29 - 0 30-39 - 9 >40 - 1			
Average Years of Experience on Admission		8.5 (3.3)				9.9 (5.5)				11.4 (5.8)				8.7 (3.6)			

1. As a classroom teacher, describe what motivated you to enroll in a Master's program.
2. Describe the extent to which those aspirations were realized.
3. To what extent did your graduate studies affect your decision to return to the classroom (or to do whatever, for those not teaching)?
4. Please describe in detail:
 a) your transition from graduate student back to classroom teaching (for students who attended full time); or
 b) the relationship between your life as graduate student and your life as classroom teacher (for those who attended part time).

 Probes (if necessary)
 - beliefs about teaching/education
 - beliefs about curriculum
 - beliefs about students and student learning
 - teaching practices
 - relationships with students
 - relationships with colleagues
5. What could your graduate program or your school do or have done to make the transition (or relationship) more positive or helpful?

Figure 1. Guiding interview questions.

than 160 km from Lethbridge, which closely approximated the geographic distribution of the study sample. However, time and distance prevented us from scheduling the final three interviews.

Interviewees within 160 km of Lethbridge were selected at random from those in each category who had provided at least a minimal response to the open-ended questions on the survey. Those at a distance were selected primarily on the basis of convenience and availability.

All interviews were semistructured. A list of guiding questions is provided in Figure 1, but rarely did the interviews follow the prescribed format completely. All interviews were conducted by one or both of the researchers at a location of the respondent's choosing. All were audiotaped and transcribed; each lasted from 30 to 60 minutes.

Detailed information about the interview sample is provided in Table 2. Of the 17 persons interviewed, eight had completed the program entirely part time and nine had attended at least one year of full-time study. Table 2 also presents information on the representativeness of the questionnaire and interview samples. It is apparent that the questionnaire sample differs little from the total available sample. Of the total available sample 45% were male and 55% were female; of the questionnaire respondents 40% were male and 60% were female. However, in the interview sample 59% were male and 41% were female. Of the available sample 43% were full time, 57% were part time. Forty-two percent of the questionnaire sample were full time, 58% were part time; 53% of the interview sample were full time and 47% were part time. Although for the most part the sample appears representative, in the interview sample there is a slight overrepresentation of full-time males and an underrepresentation of part-time females. This was due largely to availability of respondents.

Table 2
Information About Survey and Interview Sample
Relative to Total Population

	Total Population (56)		Questionnaires Returned (50)		Interviews (17)	
	Part Time	Full Time	Part Time	Full Time	Part Time	Full Time
Male	13	12	10 (20%)	10 (20%)	4 (24%)	6 (35%)
Female	19	12	19 (38%)	11 (22%)	4 (24%)	3 (18%)
Total	32	24	29 (58%)	21 (42%)	8 (47%)	9 (53%)

Analysis

Interviews and responses to open-ended questions were examined initially by highlighting and noting comments that appeared to be directly related to the research questions, or by highlighting comments, a version of which appeared in more than one interview. In many cases, where there were specific responses to particular questions (e.g., "why did you enroll in a graduate program?"), the responses were simply recorded, tallied, and categorized using representative quotes to define the category. Where responses were more lengthy and involved (e.g., in describing the relationship between graduate studies and classroom teaching), we began to identify tentative themes, recording each theme on a large sheet of paper with representative quotes. Each of these themes was supported, rejected, and/or modified as subsequent interviews were analyzed and reanalyzed. Through these constant comparative methods of "sorting and defining, defining and sorting those scraps of collected data," and organizing and coding emerging themes, eventually we were able "to place the various data clumps in a meaningful sequence" (Glesne & Peshkin, 1992, p. 133). The various themes were then clustered into three major categories, which although they are closely related to the research questions do not correspond on a one-to-one basis. These categories were motivations for enrolling in graduate studies, relationships between graduate studies and classroom realities, and the transition from graduate student to practicing professional. For the most part all themes are reported, including the numbers of respondents who espoused that particular idea and several quotations that exemplify the major idea in the theme. With respect to different transition experiences for different groups of teachers, the only differences explored in this preliminary phase were those between full-and part-time students.

Findings of the Study

Motivations for Entering the MEd Program

Questionnaire responses. The 50 persons responding to the questionnaire gave a variety of reasons for applying for graduate studies; there was little difference between the full-time and the part-time responses. The most common reason, stated by nine part-time people and 13 full-time people (44% in total; 53% of the females and 30% of the males), had to do with some aspect of personal growth, renewing energy, academic challenge, or searching for meaning.

> I wanted some time away from the classroom and school to think and to learn more about current educational theories and research ... I just needed to get rejuvenated and to get away from "the system" for a period of time and to reflect on what I did and what I wanted to do. (311, FT, F)

> I was searching for some real meaning in my life. I knew children and learning were of utmost value to me, yet I was struggling to find a sense of virtue and worth after witnessing, and experiencing, excessive destructive behaviour carried on in the name of "schooling." (314, PT, F)

The next most common reasons, stated by 37% of the females and 40% of the males, related to career advancement in some way. Eight of the part-time people and three of the full-time people (22% in total) mentioned specific job-related interests and skills that they hoped to achieve; for example: "I wanted to explore interpretive teacher-driven research by attempting to write it" (310, PT, M), and "to develop better understanding of the place of enthusiasm in learning" (354, FT, M). The second job-related category was to earn credentials to provide opportunities for career advancement or to change career direction. This particular motivator was mentioned by 20% of the sample: six of the part-time people and four of the full-time people. However, many of those 10 persons indicated that their motivation changed once they entered the program; they became less advancement-oriented and more learning-oriented. Related to these responses were those of four persons who simply stated that they had a personal goal to achieve an MEd. Two other persons indicated that they were intrigued by the concept of a "master teacher," and one other person simply said she wanted to go to school.

Interview responses. With respect to the 17 persons interviewed, the explanations for applying for graduate study became somewhat more involved. Although the major motivator was personal growth, as in the questionnaire responses, there was a difference in the gender ratio of responses. Eleven of the 17 (only three of the seven females but all but one of the 10 males, compared with 53% of females and 30% of males in the survey sample), indicated that their primary reason for applying to the program was that they "needed something for me" (5, 1) or "just wanted to develop my own potential further" (6, 1), "wanted more education about education" (11, 1), "as an avenue of getting a little bit more inside of things that are happening in education" (12, 1), and so on. Although nine of those 11 people indicated that at the time they were mostly satisfied with their teaching and simply wanted more challenge or a new stimulus, two indicated that they were currently struggling with their teaching or restless and getting stagnant. As one person expressed it,

> I think I was struggling, not in the classroom as much as out of the classroom, with teaching. I think I was doing an adequate job teaching physical education, but what I wasn't doing an adequate job with is everything else that has to do with teaching, and this is the personal relations, the staff relations, the professional growth and development. I was adequate in the classroom, but it has to go beyond that. It has to go beyond just delivering the message. (5, 1)

Even those who were satisfied indicated that they felt a "personal restlessness" (15, 1) or needed "a change from the routines" (17, 1).

Only three of the 17 interviewed stated that their primary reason for entering an MEd program was to advance to an administrative position; several others indicated that this might have been a secondary reason. Interestingly, four of the seven females but only one of the 10 males indicated that career advancement was a primary motivator. One person applied to the program because of work that he had done with a Faculty member in his school; another applied to the program because he was "frustrated, bored, upset" and "wanted to leave the pointless busy-ness of teaching" (10, 3).

Relationships Between Academic Studies and Classroom Realities

Questionnaire responses. Of the 49 persons who described the relationships between their graduate studies and their teaching, only one indicated that graduate study and classroom teaching were unrelated. Every other respondent indicated a positive relationship between academic study and classroom teaching; for example: "direct, immediate, useful, practical. Of all the PD I've undertaken, only my graduate studies changed my practice" (350, PT, F). Five of the part-time and one of the full-time respondents indicated that the primary effect of graduate study was to solidify and validate, strengthen their own beliefs, or clarify their thinking; for example: "the grad program helped solidify certain theories about pedagogy that I'd been formulating over my first decade of teaching" (310, PT). Four respondents, all part-time, spoke about the importance of the opportunity to share and meet with their colleagues. Six persons, three part-time and three full-time, mentioned direct effects on specific aspects of their teaching, such as evaluation, curriculum development, and supervision. Three part-time and five full-time persons indicated that their graduate studies had renewed their faith in the value of education or their perceptions about the importance of the classroom teacher and had brought them "closer to teaching."

However, by far the most common description of the relationships between graduate studies and teaching experiences was exemplified by nine part-time people and 12 full-time people (43% in total) who spoke about their greater understanding of the context of education, the challenge to their beliefs, their new self-awareness, and their new ability to be self critical.

> I am much more self-critical of what I do. I examine why I do things and whether or not they will truly benefit my students. If they will not benefit my students, if they are just to satisfy a curriculum requirement, or that is the way they "have to be done in this school," I do not necessarily do it or them. (311, FT, F)

> The program pushed me to think and to do things I had never dreamed of. It has given me the wisdom and confidence to promote "thinking otherwise" in my work place, to make my colleagues think of the "bigger picture." It has been one of the most powerful forces for personal and professional growth in my life. (363, FT, F)

Interview responses. All 17 persons interviewed described the relationship between their studies and their teaching in positive ways. However, there were differences between those who had been in the program full time and those who had been in the program part time. Part-time persons spoke specifically about application and about particular projects and studies they had completed. They indicated that being simultaneously a teacher and a student

"gives you that reality touch ... you can house [the academic learning] in the classroom" (12, 3). "I did a lot of testing and trying in my own classroom" (13, 3). They had also, however, become more "reflective" and "learned to think more globally" (2, 2).

> Sometimes we get hung up when we're working in the classroom and we don't have an opportunity or we don't think to reflect on why we're doing certain things, why children learn the way they do. (12, 7)

Those interviewed who had been in the program for a year or more full time commented in greater detail. They talked about their year or two at the university as being "the best years of my career" (6,4).

> It's the greatest PD I ever had been through. I'd been in the program for about a year and a half on a part-time basis ... those [courses] were marvelous; they really got me involved and it was interesting and fun and thought-provoking, I guess, to be taking graduate study courses and then going the very next day back to your classroom. There was some real advantage to that ... What I didn't realize when I got on campus [full time] was the environment of being in a university ... you just learned so much, so much more than in a class. I was really involved in a study of things. (4, 6)

> When you're there every day and you're there for two years, there's so much of what I had learned was not only in the program, it was sort of between the program, in the halls, the coffee, the golf course, you learn a lot. (10, 17)

One student, however, wondered whether he might have had additional benefits by having attended the program part time.

> Like, I still think there's maybe a benefit where teachers are actually teaching and can take courses at night and stuff. And then work in kind of slowly into their classroom, rather than doing it all and then all of a sudden going out to a completely new job. (17, 5-6)

The Transition from Graduate Studies to Teaching

Questionnaire responses. In describing the relationship between their graduate studies and their classroom teaching, questionnaire respondents touched on the notion of transition. However, the questionnaire did not elicit many responses specifically related to what it was like after the respondents were no longer graduate students. Only two respondents wrote specifically about that issue. One teacher who had been full time in the program wrote on her questionnaire: "I am finding the transition back to a classroom very challenging. Some days, fun and exciting; many days, really tough and discouraging. It seems so utterly chaotic and driven and I no longer find that appealing" (367, FT, F). Another, a male who had been part time throughout the program, also found the transition difficult.

> I found that teaching was the same old thing upon my return. In fact, some of the trivial expectations had increased. It was a let-down because at the university we dealt with philosophical aspects of education and back in the classroom, I had a job to do. I enjoyed grad studies and it was extremely refreshing, but there is very little room for higher level exploration of education in general when you are back in the classroom. (312, PT, M.)

Interview responses. This topic was explored in much more depth during the interviews. Seven of the eight interviewees who had completed their entire program on a part-time basis indicated that they had had no difficulty in returning to being a full-time professional following the completion of their master's degree. However, four of those seven persons (two males and two females) had moved into an administrative position after their studies; all four spoke of the freedom in their new positions: "Being a principal I have as much freedom as I want to take" (12, 11).

> I guess because I am an administrator, which has helped immensely, because I'm supposed to go out and evaluate teachers every year, and there you see the theory going into practice in the classroom by somebody else; and it does not become as frustrating. (1, 5)

> I teach half time as administrative assistant and half time in the library. Therefore, I see all classes one period a week. To me, I have the ideal set-up here. I see all students ... I also look after computers, so that's the area I'm interested in ... I have a lot of freedom. I have a lot of flexibility in hours which I never had in my 25 years in the classroom. (2, 6)

Another of the persons who described a relatively painless transition had continued in classroom teaching, but indicated that because of his thesis he had been contacted by the ATA and for the last eight years had been presenting workshops across the province. He had also spent one year seconded to the University as a faculty associate. Another said he felt no frustration, but found that "for the first few years after, I really tried to be creative and innovative. I'm getting more out of that now, because kids didn't seem to be able to cope" (9, 2). He indicated that he now sees much of the university teaching as ideology and the "Joe Friedmans" as reality. The last part-time student seemed to be truly content returning to the classroom. "I wouldn't do anything else" (11, 6). He considered himself a lifelong learner and indicated that he was continually learning: "you know, a 10- or 12-year personal project that I just work on" (11, 9). However, he too was applying for an administrative position.

The one part-time graduate, who admitted a great deal of frustration on her return to the classroom, had gone on to complete a PhD at another university. When she returned to her school district after two years, she was denied an administrative position and found herself back in the classroom. She "was suffering from the lack of intellectual stimulation."

> That isn't to say that I'm not really enjoying the kids or the work that I'm doing, but ... I find it very frustrating not being able to contribute and because of the work I did last year with Alberta Education, I became so aware of some of the things that were happening, and I haven't had anywhere to put that knowledge or energy or interest ... That stimulation, that thinking and not being able to run your ideas by people, you don't have that kind of thing in the classroom ... I find coming back, I try to analyze what's different, and it's moving back one step into bells and schedules that I have no control over, and, you know, that leaves me with less energy ... I'd like to be [back] in that risk-taking environment. That seems to be what I need. (3, 10-11)

Nine interviewees had completed at least one year of their MEd in full-time study; only three of those, all males, indicated that the transition back to the

classroom had not been a problem. However, of those three, one had become an assistant principal, another had accepted a position at a community college, and the third had moved to a school for special children. As expressed by the new administrator, it is "easier to change when it's totally new because you don't have the danger of falling into old habits" (4, 17). He believed that the reason he felt little frustration in returning to the classroom was not so much that he was in administration, but that his particular situation was conducive to continuing the learning that he had undertaken at the university.

> It's the first principal I've really worked with who understands in a day-to-day basis what this curriculum leadership, what it really means to be a leader in the school ... And the staff talk more about their work and activities than complaining about the kids. They talk about their teaching. (4,17)

The two others made similar comments about being in new situations where talk about teaching was common.

Six of the nine interviewees who had completed at least one year of full-time study indicated considerable difficulty with their transition back to the classroom. In fact, of those six, only four had returned to the public schools. One had moved into an administrative position, yet he still found the transition frustrating.

> I think one of the big things in schools is the loss of freedom when you come back into the system, and I use that word "system" deliberately, too. At the University there was a lot of freedom in terms of freedom of expression, academic freedom, freedom of movement, freedom from, I suppose, if I can use the term "accountability" in a sense, like nobody is keeping track of where you were, or what you were doing, or who you were with, or whatever. Once you move into a system, a school system or I suppose any other kind of system, you lose that, a lot of that autonomy. You become highly accountable to others, highly accountable to the structures of the system, the bells in the school, stuff like that ... It was a tough transition, and it still is. (15, 4)

The three who had returned to full-time teaching from full-time graduate studies, all females, described considerable frustration with their transition.

> I think I spent a lot of that year, the first year back, feeling sorry for myself, feeling that I shouldn't be where I was, that I didn't belong, feeling that the changes I had made were not accepted by my colleagues because they didn't want me to change in the first place ... And I really fought that awhile. I felt they weren't interested or accepting of my change. I was a different person, and a lot of people didn't care about that ... I felt cheated; I wasn't ready; my journey wasn't finished. (5, 6)

> I might have to dig deep to try and think of some positive things to say about it. This has been a particularly tough year. I found it just brutal ... You know, the attitudes of the general population towards education, towards teachers, and so on, continues to kind of disintegrate in a way. And so there's that. Also, you get out of the frantic pace of being in the classroom, and just readjusting to that has been hard. I don't know, in many ways the graduate school ... was so set up to treat us as professionals; you people gave us so much respect and nurturing and a lot of understanding, and your time is your own and you're very self-directed and self-motivated. When you go back to teaching, it's just ... Maybe it's the ideals that you hold dear to you while you're doing that kind of thing. You find

> back in the system that, I don't know, I just found I simply, I don't have time. I don't have time. And many of the thoughts that I had that I really cherished there, my students have not been very receptive to. And maybe that's been partly ... maybe it's me. Maybe it's the way I presented them, or my timing, or whatever. But that's made me feel kind of like withdrawing those ideas, you know. You hold them dear to you and the kids kind of poop on them, and then they're not interested. That's been discouraging for me, I would say ... I feel much more eroding of my time ... in fact, I found this a harder transition than my first year of teaching by far. I felt much more eroding of my ideals. (6, 6-7)

> I've got a feeling I'm not as content after it as I was before, and yet, you know, it was fun ... I miss just the way our minds flew a little bit beyond the here and now. I feel like I'm really in the here and now, and it's almost exhausting to try to push yourself beyond that, because you can barely keep up with the here and now. (8, 15-16)

In addition to the loss of freedom, autonomy, and opportunity these teachers experienced was the fact that their peers and colleagues seemed not to be receptive to their new ideas and in some cases even resented them.

> You go through that period of being a bit angry about all of that. You learn all kinds of new things, then you go back and you realize that there isn't a place for it. It doesn't fit. Neither do you. People don't want to hear about these new words you're using, they don't. As a matter of fact, if your administrators have not gone back to pursue their graduate degree, you are really a threat in that school. So you can learn a lot of things that affect your teaching, I think, but you can't do them. You feel that you can't do them. You feel like an outsider until you are back for about, from what I've heard anyhow, two years and then you forget a lot of what you've learned. You *have* to forget what you've learned. (7, 10)

> I felt they weren't interested or accepting of my change. I was a different person. When I went back into the classroom I was a totally different person, and a lot of people didn't care about that. I think they resent, I mean, because they see it as time off instead of time for growth, time for change. What they see is that I was out of the classroom. (5, 8)

One of these three teachers who had attempted, and failed, to enliven the staff using some of the ideas gleaned from her graduate work described herself as a "black sheep": "So, in that climate, where these teachers have been there for a long time, they sort of roll their eyes if it's anything very innovative" (8, 10).

Two of the nine interviewees, both males, who had completed full-time study found the thought of returning to the frustrations of teaching so overwhelming that they indicated that they "could not go back."

> I decided about the middle of the second semester that I quite liked this work, that I really enjoyed the kinds of things that I was doing, and realized that that would never be possible [in teaching]. I would never be able to do this kind of work and teach full time. (7, 5)

This teacher had been prepared to return to the classroom and in fact had gone to an interview at the school where he was expected to return. He described his interview: "It was unbelievable. It was like everything else I had experienced; the decisions had already been made and we were simply going

through a bit of a process to pretend as though I had some choice in the matter" (7, 7).

The second person who couldn't go back, had entered the graduate program to enliven his teaching and assumed that a master's with a focus in teaching would send him back to the classroom rejuvenated.

> What I was looking for at that point ... is I wanted to know how and why one tree was different from another tree. What the program did was to indicate that there was a whole forest and that I didn't really know that the forest existed. I thought there were just different kinds of trees, but I didn't know about the depth and breadth of it all. And at that point I realized pretty quickly that I had been sort of living a lie ... I discovered, and I think [others] discovered it too, that we simply couldn't go back to the classrooms that we had left, because we were going back as different people. We weren't going back as upgraded teachers; we weren't going back as teachers with more strategies; we weren't going back as teachers with bigger vocabularies; we weren't going back as teachers with a publication, you know, with more initials behind our names. We were going back as different human beings. (10, 5)

He described himself later as having been "contaminated" by the program, and believed that others had been as well.

> And when they go back to the school system, you know you start reading memos and dealing with wiping kids noses and getting phone calls from parents and you just want to scream: "Don't you understand what's going on here? Think about this!" ... Then you feel guilty because you maybe learned or know something that others may or may not know, but you'll never know anyway, because you never have a chance to have a conversation with them. (10, 14-15)

Discussion

Why Graduate Study

It will be no surprise to anyone that teachers undertake graduate studies for a number of reasons and that few teachers have a single motive. The findings of this study suggest a number of conclusions about why teachers pursue graduate work.

1. Teachers who enter an MEd program are for the most part excited about learning. With few exceptions students in this study considered the process, the learning, and personal growth at least as important as, if not more important than, the particular piece of paper that resulted. Fewer than one third applied because of administrative aspirations.
2. Teaching is not seen to be an affirming profession. Even those teachers who stated that they were happy with their teaching on admission to the program commented that they needed regeneration; they needed something for themselves; they were becoming stagnant; they needed a change.
3. Graduate studies appears to be one of the few legitimate ways in which teachers can take a break from classroom teaching.

Graduate Study/Teaching Relationships

With one exception, participants in this study stated that their graduate studies had had a significant and positive impact on their beliefs about teaching and their teaching practices. It was readily apparent that the teachers were enlivened by their understanding of the whys of their activities and that it was

important to them to validate their beliefs and activities. It was also important to them to have an opportunity to think about what they were doing, to reflect on it, and to understand education in a more global sense. Many teachers spoke of the personal nature of their learning, and it was clear that when classes and projects had direct application to their classroom and their teaching they were seen to be more valuable. At least three of the respondents spoke of the impact that their gender courses had had on their lives and on their teaching. Many spoke of how the program helped them to understand themselves and how important that was to them.

There were some differences between the experiences of those who had attended the program entirely part time and those who had spent at least one year of their program in full-time study. Those who had been taking courses and teaching at the same time spoke of questioning their classroom activities, of testing and retesting the ideas presented in classes, of applying what they had learned, of clarifying their thinking. Their studies were immediately applicable and relevant (or not).

All the full-time people and approximately one third of the part-time people spoke about how the program had challenged their beliefs, about their new self-awareness, their ability to be self critical, their new understandings of the context of education and how important these new insights were. Those who had been in the program full time expounded on these notions at some length and particularly valued their nonclassroom university experiences.

The following conclusions, therefore, are supported by the findings of this study.

1. Graduate studies has a positive impact on and enlivens the way teachers think about their teaching; it affords them an opportunity to share and reflect on their thoughts and ideas; it reaffirms their practices; it challenges their beliefs and expands their understandings.
2. The more the students are able to personalize their programs the greater the perceived value of the program; that is, graduates who selected courses directly related to their interests or designed assignments directly related to their classrooms found the program to be of great value.
3. Part-time study appears to lend itself more readily to a more direct questioning of specific practice and an attempt to try out new ideas in the classroom; full-time study appears to result in greater dissatisfaction with, and questioning of, existing systems.

The Transition: Graduate Student to Practicing Professional

With respect to the transition from graduate studies to once again being a full-time professional, the findings are somewhat less encouraging. Three conclusions literally seemed to leap out, particularly from the interpretations of the interviews.

1. All graduates who had returned to full-time public school teaching, whether they had been full-time or part-time students, found the transition difficult. Unless they had moved into an administrative position where they believed they had considerably greater freedom to act, or to a position with new challenges, they were no longer content with being a classroom teacher.

2. Teachers who returned to the classroom believed that their colleagues resented their return, resented being challenged, resented the new ideas that they brought to their schools. Rather than enlivening or challenging their teaching, the new ideas seemed to be simply added frustrations.
3. It appears that a full-time graduate experience is particularly detrimental to satisfaction with teaching. The challenges, the autonomy, and the opportunity for "minds to fly" during their university experience was completely stifled on the teachers' return to their classrooms.

Common (1994), however, makes a persuasive argument that exactly the opposite is the case. She argues that it is the transition from practicing professional to graduate student that has such a tremendous impact on habitual ways of living, and that it is "the change in appellation from practitioner to student [that] represents a forfeiture of essential powers: the power to define self, to interpret context, to determine action, and to judge worth" (p. 15). Common supports her position by describing how truth is constructed in academe "within an epistemological paradigm having firm traditions and established knowledge" (p. 19), and how that march to knowledge is dated, ponderous, and prescriptive, in contrast to the professional paradigm where practical and professional knowledge is valued and change is ever-present.

However, the professional paradigm Common (1994) describes appears more typical of business, law, medicine, and engineering than it does of teaching. We would also suggest, with an obvious bias, that the academic paradigm Common describes does not characterize the University of Lethbridge Master of Education program. Therefore, we remain convinced that the following conclusions are warranted.

1. Full-time university graduate studies are relevant to teaching only in the way that a vacation is; that is, they are exciting, stimulating, and refreshing, but after a few weeks back home all that's left are the photographs; and/or
2. Schools are damaging many teachers' well-being. In order to thrive, most teachers need an opportunity to grow, to risk, to be challenged intellectually; not only do some schools not facilitate this development, they actively prevent it for at least some teachers.

These are damaging statements. One might suggest with respect to the first that this sample is atypical and that the findings would not generalize to other graduate programs. We would argue, however, that if these conclusions are true of the University of Lethbridge Master of Education program they may be even more true of other Master of Education programs. the University of Lethbridge program was designed specifically with the classroom teacher in mind and was intended to promote the concept of the master teacher. Graduates are expected to return to the classroom, and students select courses and design course assignments to complement and enhance their classroom activity. The four core courses challenge students to think more critically and broadly in the areas of curriculum, research, educational practices, and teaching and teacher development. Therefore, it is likely that if the graduates of this program feel alienated from the classroom, graduates of other programs are at least as likely to do so. This premise is being tested as the study is expanded to include the other Alberta universities.

With respect to the second observation, this study merely verifies in a particularly obvious manner what many researchers and educators have been saying since Lortie (1975) so eloquently described the issue in his book *School Teacher*.

Recommendations

The sample for this study was limited to the graduates of one university's MEd program, and the first recommendation clearly must be to further analyze the data with respect to gender, age, and experience and to expand the study to graduates of other universities. Nevertheless, our experiences and the findings of this study suggest, at least to us, some ideas that are at least worth considering.

We would suggest first, with respect to MEd programs, that:

1. *University faculty members should be visible in schools and need to use schools as a basis for their research.* This kind of activity not only encourages teachers to become graduate students and to further their education, it also allows them to maintain their excitement and continue their involvement with learning and with the university. As two graduates expressed it in describing situations in which a faculty member was doing research in their schools:

 > I really felt that although he never called it research, you felt that you were involved in generating new things in your classroom, and that enlivened the teaching, it really did ... You cannot do it by yourself, so there needs to be some sort of action research going on in school. (7, 7-8)

 > It's the forest and the trees. When we're teaching, we're right in the trees and I think the university is able to view the whole forest and I think it's important that we're able to get a view of the forest, because otherwise we'll never get out of those trees, you know. It's really important that the university keep doing this. (8, 11)

2. *Master of Education programs need to be made personal.* It is important to target master's programs to teachers' needs and to allow teachers the flexibility to identify those needs and select their courses and their activities accordingly, and as much as possible to carry out assignments and projects in actual classrooms, preferably their own. As Clark (1992) puts it, on the basis of more than 10 years of research on teacher thinking, "Experienced teachers can become designers of their own personal programmes of self-directed professional development" (p. 75).

 However, it would also appear to be important that certain courses and/or experiences be required. Almost all the graduates in this study spoke of the importance of the core. Many indicated that they would not have chosen those courses had they not been required, but that those courses were invaluable for expanding their horizons, for making them see themselves, teaching, students, and curriculum in different ways. The challenge for graduate programs will be to find the balance between personalizing each program and requiring common experiences.

3. *Master of Education program faculty should address directly with their students the issues they are likely to face when they return to their classrooms.* This suggestions did not arise from the data; rather, it was made by a number of

participants who read the study and by current students with whom the study was discussed.

Finally, we cannot resist making a couple of suggestions to schools and to teachers. The data from this study clearly identified schools as places that inhibit many teachers' growth and creativity and many teachers as less than receptive to new ideas and challenges. Therefore, we suggest that:

1. *Teachers need to be for each other.* Sarason (1993) points out that "it is paradoxical to criticize teachers and then to come up with guidelines asking teachers to change themselves" (p. 268). Nevertheless, teachers are professionals and must assume some individual responsibility for their professional development. It is not always necessary to make big changes or dream big dreams; often change begins with a small group of people trying to make a difference. This fact is well illustrated by the following graduate:

 > There are things that schools could do ... I've really come to believe that the teachers are the school. It's just that they don't think they are ... They were convinced that there wasn't time to do any of this and I simply went out there and encouraged them once in a while. We started once every two weeks to meet for an hour to talk about what we were doing, you know, and as the year went on they began to realize that you just have to say, well, after school on Wednesday we're going to do this for an hour, period ... So it was a very, very small thin, it wasn't big ... and it worked. (7, 8-9)

 One simple strategy might be to ensure that at least two teachers from a school attend graduate studies together. Much more is possible with the support of even one other person.
2. *The entire structure of schooling needs to be changed to take into account the needs of teachers to grow, to learn, and to live.* This kind of change requires exciting, innovative, and courageous leadership. Schools should be structured to take advantage of the expertise of teachers, to create opportunities for their professional development, to respect their learning, to build in time for them to continue learning, and to give their learning legitimacy. The implication here is that we need to change to way we think about teachers and teaching. Burger (1994) describes a large study in which "teachers with stories of growth said the major impetus for their continued development had been the school culture and practices which emphasized that everyone in the school was a member of a learning community" (p. 3). Clearly this is no small task and is well beyond the scope of this small study, but the data from this study support the notion that this type of change is long overdue.

References

Baskett, H.K.M., Marsick, V.M. and Cervero, R.M. (1992). Putting theory to practice and practice to theory. *New Directions for Adult and Continuing Education, 55,* 109-118.

Burger, J. (1994, July). *The case for substantial reform of teacher education.* Paper presented at the Canadian Society for Studies in Education Conference, Calgary.

Clark, C.M. (1992). Teachers as designers in self-directed professional development. In A. Hargreaves & M. Fullan (Eds.), *Understanding teacher development.* New York: Teachers College Press.

Common, D.L. (1994, July). *Still waters run deep: Professionals as returning graduate students.* Paper presented at the Canadian Society for Studies in Education Conference, Calgary.

Duffy, G., & Roehler, L. (1986). Constraints on teacher change. *Journal of Teacher Education, 37*(1), 55-58.

Glesne, C., & Peshkin, A. (1992). *Becoming qualitative researchers: An introduction.* New York: Longman.
Greene, M., & Campbell, C. (1993). *Becoming a teacher: The contributions of teacher education.* Edmonton, AB: Alberta Department of Education.
Hargreaves, A., & Fullan, M. (1992). *Understanding teacher development.* New York: Teachers College Press.
Lortie, D. (1975). *Schoolteacher: A sociological study.* Chicago, IL: University of Chicago Press.
McNergney, R.F., & Carrier, C.A. (1981). *Teacher development.* New York: Macmillan.
Odell, S.J. (1986). Induction support of new teachers: A functional approach. *Journal of Teacher Education, 37,* 26-29.
Sarason, S.B. (1993). *The case for change: Rethinking the preparation of educators.* San Francisco, CA: Jossey-Bass.

The Alberta Journal of Educational Research *Vol. XLI, No. 2, June 1995, 231-235*

Book Review
Women's Work is Never Done: Resolving Gender Inequalities in Education

Gendered Education: Sociological Reflections on Women, Teaching and Feminism. Sandra Acker. Toronto: Ontario Institute for Studies in Education Press, 1994, 193 pages.

Reviewed by *Diane G. Symbaluk, University of Alberta*

Few university students today would be surprised to find themselves in a science or medical course taught by a female instructor with predominantly female enrollment. Clearly the same cannot be said for the 1960s, 1970s, or even early 1980s when female professors were scarce and the female student minority registered mainly in sex-stereotyped courses. Women were encouraged to take courses in elementary education or family studies and were discouraged from enrolling in masculine subjects like engineering or business management. Despite the fact that women today enjoy more academic freedom in what Fulton (1993) describes as the "decade of the woman student," double standards persist and prevent women from achieving equal status with men in virtually all spheres of life.

Gendered Education examines women's progress in education over the last 30 years. The emphasis is on the continuity of women's inferior status as students and academics despite advances procured since the second wave of feminism in the 1960s. Acker recalls her own experiences and the political climate of the late 1960s and early 1970s that contributed to her research on gender issues. These recollections help the reader understand why women continue to have low academic aspirations and why they obtain so few doctorates relative to men. Critical themes of the book include the neglect of women in higher education, advances in feminist theory and research, and the implications of gender biases for teaching and women's advancement in academics.

Acker's account begins with a personal description of life as an undergraduate and graduate student in the female-resistant educational system of Britain at that time. My interest was piqued by parallels I was able to draw between Acker's experience in Britain and my own experience as a female graduate student in Canada some 20 years later.

I was impressed by Acker's recollection of her aspirations as a new graduate student in sociology as well as her awareness of how her views changed with her progress in the educational system (Chapter 1). Acker describes the discovery of sociology as "a revelation ... a means of making the world better for its less fortunate inhabitants." The revelation inspired her to go on to pursue a master's degree in education studying the socially disadvantaged. Acker explains how her experience altered her plans to return home to teach after obtaining a master's degree:

> My original ambition was completely outside the department's preferences and snobberies, into which I was gradually induced: greater prestige derived from theorizing about educational systems and abstract schools than from going into real ones ... it was better to do a doctorate and enter academic life than to stop with the delightfully dubbed "terminal masters." (p. 13)

Accepting the inevitable, Acker went on to obtain a doctorate, researching how other graduate students absorbed the same "academic ideology." Distinct from the original focus of her analysis, Acker discovered a common theme in women's interview responses that was not evident in statements made by men. Specifically, women had relatively low academic aspirations and seldom viewed themselves outside of sex-stereotyped career ambitions (e.g., elementary school teacher, nurse, librarian).

Acker holds the educational system responsible for gender-biased views in its blatant neglect of women in the sociology of education prior to the 1980s. This issue is covered in detail in Chapter 2, which is appropriately titled "No Woman's Land." The chapter includes highlights from Acker's (1981) research on gender issues in major sociological journals over a 20-year period. Because men are more likely to occupy high-prestige positions in society, their work is considered important whereas women's is viewed as inconsequential. As a result of these assumptions, researchers have studied male-dominated professions almost exclusively.

> Researchers could have studied the educational and occupational socialization of hairdressers just as easily as that of printing apprentices. Or the social and educational backgrounds of nurses or social workers or actresses. Or the workings of university departments of French or technical college departments of catering. Or the role of further education in mobility chances of working-class girls. But no-one did. (p. 31)

One implication of male-biased research sampling is that findings for studies based on all-male samples may not be the same as results obtained using women. Acker claims this point is ignored by most researchers in education, who either exclude women from their research samples altogether or fail to interpret findings for female samples when they are included along with results obtained for males.

Feminist writers in other disciplines have also pointed to problems associated with male-biased sampling. For example, feminists have criticized criminology theory and research for its use of all-male samples (Naffine, 1987), and overreliance on concepts and results that are gender-specific (Chesney-Lind, 1989). For example, items used to measure delinquency (e.g., number of acts of assault or vandalism) are more characteristic of males than females. In contrast, girls commit more status offenses (e.g., running away from home or curfew violation) but these behaviors are not considered delinquent acts.

Although feminist ideology is starting to permeate criminology, Acker notes that gender issues are still not an integral part of the sociology of education. Instead, gender issues remain almost the exclusive concern of women in education, and more specifically feminist scholars. Importantly, Acker points out how gender issues are problematic even for female academics who study them.

> It is just as difficult for a feminist academic to know to what extent the reception accorded to her work or the response to her teaching is related to her politics, her gender, or some other feature of her background, demeanor or ability, or whether is resides in the politics and practices of the institution, or even the culture and political economy of the society. (p. 70).

Acker claims that a feminist approach will increase the salience of gender issues and may resolve some of the problems associated with biased sampling and sex differences in educational research. She provides a list of "tasks" for a feminist approach to the sociology of education. One is to reinterpret previous research (where possible) to include a discussion of results on women and an explanation for any inconsistencies in findings for females relative to males.

Chapters 3 and 4 outline the major contributions and limitations of various feminist approaches to gender issues and discuss some of the current debates. One issue is whether research by a feminist scholar always constitutes feminist research. In this regard Acker points to the dangers of rigid inclusion criteria claiming that "as feminist scholarship becomes increasingly institutionalized, there is a tendency for it to develop its own rituals and practices of exclusion" (p. 69). She suggests that feminist scholarship would benefit more from an assessment of what such research contributes to the advancement of knowledge than from a concern with what does nor does not constitute feminist research.

The second part of this book focuses on the marginality of women as teachers. The chapters in this section examine why teaching is viewed as a "semi-profession" (Chapter 5), why the educational system is reluctant to adopt "anti-sexist initiatives" (Chapter 6), and how gender is implicated in the career structures of teachers (Chapter 7). Acker suggests that teaching is considered a semi-profession because it is dominated by women who are still viewed as the subordinate sex.

Acker clarifies this contention in Chapter 5, providing evidence that female-based majorities are largely restricted to elementary and junior grades with women becoming increasingly scarce as the age of students increases or the prestige of the teaching position increases. Although a great deal of sociological research is directed at the implications of gender for education and teacher careers, Acker suggests that much of this literature suffers from major flaws including an overreliance on blame-the-victim approaches and an inability to view women outside of familial roles: "Such approaches are one-sided, placing the emphasis on the individual woman as 'actor' with little or no attempt to assess the structures within which action takes place" (p. 87).

In other words, few studies examine how the social structure of education determines women's roles as teachers. For example, Acker claims that research on female teachers' favoritism toward male students would benefit from an analysis of effects of colleague and department heads' expectations on teachers' attitudes and behaviors.

In addition to shortcomings in research on gender issues, Acker claims that even when "gender-equal initiatives" exist, teachers are reluctant to incorporate them into their classrooms (Chapter 6). She goes on to outline four factors that contribute to this resistance including implementation issues, teacher demographics, teacher ideology, and teacher work conditions. Im-

plementation issues, for example, may consist of a lack of awareness on the part of individual teachers who seldom read academic journals that contain gender-neutral initiatives; it can also include a lack of communication between teachers with similar interests in the same school or from other schools. Communication of gender-neutral practices is an important consideration and "the cost of implementing anti-sexist innovations will decrease when it becomes clearer what they are and how one goes about putting them in practice" (p. 103).

Chapter 7 describes teacher career structures in England and Wales for 1990. The purpose of this chapter is to outline teaching as a career from both social-structural (macro) and individual (micro) perspectives. Using a macro-sociological perspective, Acker provides a breakdown of the percentage of teachers and teaching heads who are women as well as the distribution of male and female teachers by grade level. Through interviews with teachers in two primary schools she also provides a micro-sociological analysis of teaching based on subjective interpretations of the profession as well as expectations and feelings of marginality among female teachers.

Part 3 focuses on the professional experiences of female academics who are often "double-burdened" with career and family commitments. In addition, these women hold a "minority position" among faculty members, and this marginal status has implications for achievement, role responsibilities, and interactions with other colleagues.

For example, Chapter 8 points out that women's academic achievements are thwarted largely as a function of time restraints resulting from unsuccessful attempts to juggle research commitments with extracareer responsibilities related to raising a family. Chapter 9 provides an in-depth analysis of the status and role of female academics in British universities. This chapter examines explanations for women's low academic standing (e.g., early socialization and conflicting role responsibilities) and includes strategies for improvement that follow from radical and liberal feminist approaches.

Acker concludes by noting how some aspects of the educational system have undergone drastic changes over the last 20 years, but these changes were slow to occur and many years passed before improvements were detected. Consequently, many gender issues are still not resolved (Chapter 10). There are far greater numbers of female academics in universities throughout the world and women are increasingly enrolling in traditionally male-dominated courses such as engineering or business management. Importantly, many of these women have been awarded research grants or scholarships for outstanding work in their fields, and some have gone on to pursue doctorates. As a result, we are seeing more female graduate students and more female professors.

In spite of these considerations, the number of tenured female academics still lags. Similarly, although many women are publishing their research in journals, relatively few are invited to write edited collections. I question Acker's suggestion that these observations are signs that qualified women are still being excluded from academic circles, and would be more apt to conclude that women still have lower aspirations relative to men, with fewer women seeking to obtain tenure or expressing an interest in publishing edited books. Despite educational advances in the form of acknowledgment for work or

policies concerning equal pay, female academics are still both at an economic and status disadvantage relative to men.

I commend Acker for addressing issues that are especially problematic for female academics including the timing of marriage and families. The decision to have children has implications for the academic careers of female students in their childbearing years. Academic choices will in part be influenced by commitments to motherhood and may include postponements or interruptions in programs as well as lowered educational aspirations (Chapter 7). Although female academics beyond their childbearing years appear free to pursue academic careers, they often have outside employment or children at home. In reality these women are overburdened with responsibilities including domestic tasks that impede their professional careers (Chapter 8).

Although recognizing that women may opt not to have children early on, Acker assumes these women are merely postponing childbirth until they have completed their postsecondary education. This assumption is evident in her statement that childbearing will occur at a time when expectations for occupational performance are high (Chapter 8). The author leaves out one category of female academics, those who forgo childbirth altogether in order to obtain a degree and secure an academic career in a highly competitive job market.

I contend that childless women, as a subgroup of the female population, may experience greater marginality than women in general. Greater marginality occurs because childless women (of childbearing years) break with traditional norms for women in our society; women are expected to marry and become mothers shortly thereafter. These assumptions can have implications both for securing a tenure-track position at a prestigious university and for exacerbating marginality in relations with colleagues. Employers may be reluctant to hire childless women because they anticipate interruptions as a result of childbirth. Also, childless women do not share common experiences with men or with women who are mothers, and this issue merits further exploration.

Gendered Education is an interesting, witty, and educational look at gender issues. Although materials included in this book are largely based on studies conducted in the United Kingdom or reported in British journals, the information is representative of what occurs in other industrialized countries and would, therefore, be of interest to people outside Britain. A variety of audiences can benefit from reading this book, but it is of more immediate relevance to postsecondary students and academics in sociology, education, or women's studies.

References

Acker, S. (1981). No-woman's-land: British sociology of education 1960-1979. *Sociological Review, 29,* 77-104.

Chesney-Lind, M. (1989). Girls' crime and women's place: Towards a feminist model of female delinquency. *Crime and Delinquency, 35,* 5-29.

Fulton, O. (1993). Women catch up. *The Times Higher Synthesis: Higher Education Trends, 16,* ii-iii.

Naffine, N. (1987). *Female crime: The construction of women in criminology.* Boston, MA: Allen and Unwin.

Diane Symbaluk is a doctoral student in the Department of Sociology. Areas of specialization include social psychology, criminology, and research methodology.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
845 Education Centre South
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1438

ajer

THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XLI NUMBER 3 SEPTEMBER 1995

Special Issue
Canadian Perspectives on *The Bell Curve*

The Alberta Journal of Educational Research

Published in March, June, September, and December by the Faculty of Education, University of Alberta.

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane

Consulting Editors

Terry Belke
Mount Allison University

George Buck
University of Alberta

Ardra L. Cole
Ontario Institute for Studies in Education

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Heather Ryan
Massey University

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Robert H. Short
Fern Snart
Ken Ward

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, the Alberta Advisory Committee for Educational Studies, and the Research Grants Office of the University of Alberta

The subscription rate is $37.00 per year for individuals, $45.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Office of the Dean
845 Education Centre South
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0236

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XLI, Number 3, September 1995

Special Issue: Canadian Perspectives on *The Bell Curve*

Judy Cameron **237** Introduction

Overview of *The Bell Curve*

Terry W. Belke **238** A Synopsis of Herrnstein and Murray's *The Bell Curve: Intelligence and Class Structure in American Life* (New York, Free Press, 1994)

Commentaries on *The Bell Curve*

Scientific, Methodological, and Statistical Critiques

Douglas Wahlsten **257** Increasing the Raw Intelligence of a Nation is Constrained by Ignorance, Not its Citizens' Genes

Raymond E. Fancher **265** *The Bell Curve* on Separated Twins

David John Bateson **271** How Can *The Bell Curve* Be Taken Seriously?

Parmeswara Krishnan **274** *The Bell Curve*: Some Statistical Concerns

John Gartrell
Stephanie Amadeo Marquez **277** The Spurious Relationship Between IQ and Social Behavior: Ethnic Abuse, Gender Ignorance, and Confounded Education

Concepts of Intelligence and IQ

Linda S. Siegel **283** Does the IQ God Exist?

Robin Barrow **289** Keep Them Bells a-Tolling

Reflections

Moya Kavanagh **297** *The Bell Curve*: Rising Fears, Fallen Dreams

W. Frank Epling **302** The Bell Bottom

Eric D. MacPherson **304** *The Time Machine*: The Next Generation

Social Issues

Rebecca Priegert Coulter **308** And Academic Sexism Too: A Comment on *The Bell Curve*

Ruth Reviere **312** "Slow Through The Dark": A Commentary on *The Bell Curve*

Erin Gibbs Van Brunschot Augustine Brannigan **316** IQ and Crime: Dull Behavior and/or Misspecified Theory?

Policy Implications

John R. Kirby **322** Intelligence and Social Policy

David W. Livingstone **335** For Whom *The Bell Curve* Toils

John B. Connors **342** Two Tails are Better than One: The Logic of *The Bell Curve*

Educational Challenges

Catherine D. Krull W. David Pierce **349** IQ Testing in America: A Victim of its Own Success

Michael Schleifer **355** Should We Change our Views About Early Childhood Education?

David G. Wangler **360** Is *The Bell Curve* a Ringer?

David F. Robitaille Edward C. Robeck **367** Distorted Vision: Education as Seen Through *The Bell Curve*

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 237-237*

Special Issue:
Canadian Perspectives on The Bell Curve

Introduction

I am pleased to introduce a special issue of the *Alberta Journal of Educational Research* that deals with Richard Herrnstein and Charles Murray's (1994) controversial book *The Bell Curve*. Following the publication of *The Bell Curve*, many commentaries, critiques, and reviews have appeared. There has even been a set of readings edited by Jacoby and Glauberman (1995) that is almost as lengthy as Herrnstein and Murray's book itself. What makes the present collection different is that it portrays the perspectives of Canadian academics.

This series of readings begins with an excellent overview of *The Bell Curve* that was written by Terry Belke, presently an assistant professor at Mount Allison University. Dr. Belke recently received his PhD in psychology from Harvard where he took graduate courses from and participated in research with Richard Herrnstein. In my opinion, Belke's paper provides a fair and accurate overview of the book.

The remaining papers are insightful commentaries on *The Bell Curve* written by psychologists, sociologists, and educators from across Canada. The commentaries range from questions concerning the appropriateness of the statistics to philosophical essays. The issues that are raised in the book and in the commentaries in this special issue are important, have a potentially significant impact for social policy and education, and will be discussed for years to come.

Judy Cameron
Editor

References

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Jacoby, R., & Glauberman, N. (Eds.). (1995). *The bell curve debate.* New York: Times Books.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 238-256*

Terry W. Belke
Mount Allison University

A Synopsis of Herrnstein and Murray's *The Bell Curve: Intelligence and Class Structure in American Life* (New York, Free Press, 1994)

The Bell Curve: Intelligence and Class Structure in American Life by Richard J. Herrnstein and Charles Murray has been the source of considerable debate in recent times. Given the controversial nature of the subject, *The Bell Curve* deserves a careful and considerate reading. Readers will bring to their reading of this book an initial attitude toward the subject that will color their interpretation of the message that Herrnstein and Murray are trying to convey. The reader who is vehemently opposed to any claim that "intelligence" has a genetic basis will dismiss the words as rantings of racists. At the other end of the spectrum, readers consumed with hatred and bigotry toward individuals of other races and ethnicity will interpret the contents of this book as justification for their biases and actions. In either case, these readers will be doing this work a disservice. To understand what this book has to offer and to keep it in perspective, the reader should endeavor to suspend these initial attitudes so that the book can be read without prejudice. This is not to say that the analyses, interpretations, and forecasts offered by Herrnstein and Murray are beyond critique, but that they should be heard.

The introduction begins with an overview of the history of the concept of intelligence beginning with Galton and tracing the developments surrounding the IQ controversy to modern times. Galton first used tests of perceptual and motor skills to discriminate differences in intelligence. Later Binet developed tests of reasoning, drawing, analogies, and pattern recognition that form the basis of modern intelligence tests. Spearman's contribution was the concept of a general intelligence factor (g) underlying correlations between tests of intelligence. Early advances in the study of intelligence were reversed by advocacy of testing for racial policies (e.g., sterilization laws). Finally, the 1960s heralded a fundamental shift away from causes within the individual as the source of social ills to causes outside the individual. Social factors that could be redressed by the government were considered the source of deficiencies. In this context of egalitarianism, recognition of biological bases of individual differences was and remains anathema.

The introduction closes with an outline of Herrnstein and Murray's perspective. First, the authors warn the readers against committing the ecological fallacy, that is, generalization from the aggregate to the individual, given that the analyses to be presented are for aggregate data. Second, the authors state that the importance of intelligence among human virtues has been inflated and

Terry Belke received his doctorate from Harvard University in 1993. As a graduate student in the Department of Psychology he worked with Richard Herrnstein from 1988 to 1991.

that the assumption that a person's intelligence can be inferred from casual interactions is erroneous. Third, that the identification of IQ with attractive human qualities is wrong. Most importantly, readers should keep the following statement in mind as they read the book.

> Measures of intelligence have reliable statistical relationships with important social phenomena, but they are a limited tool for deciding what to make of any given individual. Repeat it we must, for one of the problems of writing about intelligence is how to remind readers often enough how little an IQ score tells you about whether the human being next to you is someone whom you will admire or cherish. (p. 21)

Part I: The Emergence of a Cognitive Elite

This part documents the transformation of society during the 20th century from one in which social standing was largely determined by birth and cognitive ability was evenly spread throughout social strata to a society in which social mobility is a function of cognitive ability.

In the chapter entitled "Cognitive Class and Education, 1900-1990," Herrnstein and Murray document the opening of colleges to the general population that led the way to the cognitive partitioning of the American population. Three processes were initiated: the growth of the college population, a more efficient recruitment of cognitive ability, and further sorting of cognitive ability by the colleges. From 1900 to 1990, the percentage of 23-year-olds with bachelor's degrees increased from 2% to 30%. However, this increase in the probability of going to college was not evenly spread across the range of cognitive ability. As the authors show, the probability of entering college increased dramatically for students in the upper half of the IQ distribution, but decreased marginally for those in the lower half. Finally, this influx of cognitive ability into colleges was not evenly distributed among colleges. Colleges of greater prestige harvested a greater yield of top students than did colleges of lesser prestige. This process was aided by long-distance travel becoming commonplace and an increase in the number of families that could afford to send their offspring to elite colleges. They close this chapter with a discussion of the isolation of different segments of society created by cognitive partitions.

In the chapter entitled "Cognitive Partitioning by Occupation," Herrnstein and Murray document a second process of occupational sorting by cognitive ability. High-IQ professions have grown tremendously since 1940, and the proportion of individuals in the top decile of IQ in these professions has increased. Another line of evidence for occupational cognitive segregation offered by the authors is the decline in CEOs with only high school and the concomitant increase in CEOs with graduate degrees. The point of this chapter is that

> in midcentury, America was still a society in which a large proportion of the top tenth of IQ, probably a majority, were scattered throughout the population, not working in a high-IQ profession and not in a managerial position. As the century draws to a close, a very high proportion of that same group is concentrated within those highly screened jobs. (p. 61)

In "Economic Pressures to Partition," the third chapter, the authors argue that worker productivity is directly linked to intelligence and that this rela-

tionship is sufficient to have economic consequences. This thesis directly contradicts the received wisdom that "the relation between IQ scores and job performance is weak, and, second, whatever weak relationship there is depends not on general intellectual capacity but on the particular mental capacities or skills required by a particular job" (p. 66). Herrnstein and Murray suggest that this received wisdom is repudiated by meta-analyses that show that job performance is well predicted by broadly based tests of intelligence, that the general intelligence factor is a better predictor of job performance than tests of specific skills, and that correlations between tested intelligence and job performance are higher than previously considered (p. 70). Tests of cognitive ability have been shown to be better predictors of job performance ratings than biographical data, reference checks, education, interview, college grades, interest, or age. In terms of economic consequences, Herrnstein and Murray illustrate through example the difference in the value of the productivity of an employee at the 50th and the 84th percentile of the IQ distribution. The magnitude of the difference in productivity varies with the complexity of the job. Estimates of the costs entailed in disallowing hiring based on intelligence range from a maximum loss of 80 billion to a minimum loss of 13 billion. The main point of the chapter is that intelligence is directly related to job performance and that "getting rid of intelligence tests in hiring—as policy is trying to do—will not get rid of the importance of intelligence" (p. 88).

In "Steeper Ladders, Narrower Gates," Herrnstein and Murray attempt to forecast the social consequences of continued cognitive partitioning in education and occupations. First of all, the value of intelligence in the marketplace has increased. Wages for individuals in high-IQ occupations have grown more rapidly than the salaries of low IQ occupations. According to the authors, this increasing disparity in wages reflects an increasing economic demand for intelligence. "The more complex a society becomes, the more valuable are the people who are especially good at dealing with complexity" (p. 99). Along with this cognitive partitioning comes segregation of individuals of different cognitive abilities in both the workplace and community. Finally, Herrnstein and Murray discuss the implications of the conclusion that success in life will be based to some extent on inherited differences in cognitive ability among people (p. 108). As the social environment becomes more uniform, the proportion of variance attributable to inherited differences in cognitive ability increases; therefore, success in a society stratified by cognitive ability depends increasingly on inherited differences and decreasingly on the social environment (pp. 109-110). In addition, Herrnstein and Murray suggest that there may be an increasing tendency for individuals of similar levels of cognitive ability to marry aided by the feminist revolution. In summary, Herrnstein and Murray state that three phenomena are the result of these sorting processes:

1. The cognitive elite is getting richer, in an era when everybody else is having to struggle to stay even.
2. The cognitive elite is increasingly segregated physically from everyone else, in both the workplace and the neighborhood.
3. The cognitive elite is increasingly likely to intermarry. (p. 114)

Part II: Cognitive Classes and Social Behavior

Part II documents the relationship of cognitive ability to social behaviors. According to the authors, "high cognitive ability is generally associated with socially desirable behaviors, low cognitive ability with socially undesirable ones" (p. 117). In this part, Herrnstein and Murray argue that intelligence rather than socioeconomic status is responsible for group differences in social behavior. Furthermore, they maintain that the potential causal role of cognitive ability has been neglected and that even if cognitive ability is not the cause of differences in social behaviors, viewing the problem from the perspective of cognitive ability will contribute to our understanding of these social problems. The analyses in this part were done using the National Longitudinal Survey of Labor Market Experience of Youth (NLSY). All analyses were conducted with non-Latino white individuals to demonstrate that the relationship of cognitive ability to social behavior is independent of race and ethnicity (p. 125).

In Chapter 5, entitled "Poverty," Herrnstein and Murray show that low intelligence is a stronger precursor of poverty than coming from a low socioeconomic status background. The proportion of Americans living below the poverty line decreased from 50% in 1940 to 15% in 1970 and has remained around 15% since 1970. Using the NLSY, the percentage of white youths living in poverty is shown broken down by parents' socioeconomic status (SES) and by cognitive class of the individual. In both analyses, the percentage of individuals living in poverty increases with decreasing parental SES and decreasing cognitive ability. When both parental SES and intelligence are placed in the same analysis, the relationship between IQ and probability of being in poverty is stronger than the relationship between parental SES and the probability of being impoverished. Herrnstein and Murray conclude that "cognitive ability is more important than parental SES in determining poverty" (p. 135).

When the relationships are further broken down into educational categories, the relationship between IQ and poverty is stronger than the relationship between parental SES and poverty for individuals with a high school diploma. However, for individuals with a bachelor's degree, neither IQ nor parental SES shows a relationship with poverty. For children, the probability of living in poverty is higher the lower the IQ of the mother, independent of marital status. When marital status is considered, the relationship between mother's IQ and the probability that a child will live in poverty is stronger for mothers who are separated, divorced, or never married than for married women. The relationship between parental SES for mothers and the probability that a child will live in poverty is weaker. In conclusion, the authors state that "the high rates of poverty that afflict certain segments of the white population are determined more by intelligence than by socioeconomic background" (p. 141).

In the chapter entitled "Schooling," according to Herrnstein and Murray dropping out of school is a recent phenomenon that developed with "the assumption that it is normal to remain in school through age 17" (p. 144). The proportion of people who obtain a high school diploma rose throughout the century from fewer than 10% in 1900 to around 80% in 1964 and then leveled off at around 75%. Early in the century the IQ gap between those who dropped out and those who completed high school was marginal; however, since the 1950s the gap between those who complete and those who do not increased to

approximately 15 points. Analysis of the NLSY showed that the percentage of students who dropped out were almost entirely from the bottom quartile of the IQ distribution. Furthermore, the relationship between probability of dropping out and IQ score was stronger than the relationship with parental SES. For temporary dropouts, the relationship between parental SES and the probability of getting a GED (General Educational Development) instead of a high school diploma was stronger than the relationship with IQ. Conversely, the probability of getting a bachelor's degrees increases with IQ and the relationship with IQ is stronger than the relationship with parental SES. In sum, both socioeconomic background and cognitive ability play a role in performance in school.

In the chapter "Unemployment, Idleness, and Injury," analyses of the NLSY data by Herrnstein and Murray showed that the probability of being out of the labor force for a month or more decreased as IQ scores increased but increased as parental SES increased. Low cognitive ability also increased the risk of being off work due to disability. In conclusion, Herrnstein and Murray qualify these relations by noting that "most white males at every level of cognitive ability were in the labor force and working, even at the lowest cognitive levels" (p. 165).

"Family Matters," the next chapter, explores the relationship of cognitive ability to marriage, divorce, and illegitimacy. Analyses of the NLSY confirm the general view that the average age at first marriage tends to be high for individuals higher in cognitive ability than for those lower in cognitive ability. In terms of the probability of marriage by age 30, for individuals with a high school educational level the probability increases with IQ and declines marginally as parental SES goes from low to high. For individuals with college, the probability does not vary with either parental SES or IQ. In terms of divorce, analysis of the NLSY showed that the probability of divorce within the first five years of marriage decreased as IQ went from low to high, but increased as parental SES went from low to high. In terms of educational levels, for individuals with a college education, the probability of divorce decreased as IQ scores increased. No relation was observed for high school graduates. The rate of illegitimate births has increased markedly in the past 30 years from 11% in the 1960s to 30% in the 1990s. Analyses of the NLSY showed that the percentage of white women who had ever given birth to an illegitimate child increased from 2% to 32% as cognitive class moved from very bright to very dull. For both IQ and parental SES, as values went from low to high, the probability of an illegitimate birth declined; however, the relationship of IQ to illegitimacy was stronger than the relationship of parental SES. Other analyses showed that the odds of a child being born out of wedlock was lowest if either the mother or both parents were absent by age 14 and highest if only the father was absent. In addition, the probability of the first child being born out of wedlock for white mothers living in poverty decreased markedly as IQ went from low to high and increased markedly as parental SES went from low to high. Herrnstein and Murray conclude that "low intelligence is an important independent cause of illegitimacy" (p. 189). Furthermore, comparing the top and bottom quartiles of IQ, the authors note that in the top quartile of cognitive ability, the percentage of households that consist of a married couple is higher whereas the percentage

of households that have experienced divorce and percentage of children born out of wedlock are lower than for the bottom quartile of cognitive ability.

In "Welfare Dependency," Herrnstein and Murray showed that the probability of mothers going on Aid to Families with Dependent Children within a year of first birth decreased as both IQ and parental SES went from low to high with a stronger relationship found for IQ. When the analysis focuses only on chronic welfare recipients rather than all welfare recipients, the probability of a white woman becoming a chronic welfare recipient decreases as both IQ and parental SES go from low to high; however, in this case, the relationship with parental SES is stronger. Education played a role in that women with high school or less were at risk for becoming chronic welfare recipients, not women with more than high school. Herrnstein and Murray conclude that "having a baby without a husband is a dumb thing to do. Going on welfare is an even dumber thing to do, if you can possibly avoid it" (p. 200).

In "Parenting," Herrnstein and Murray ask, "is the competence of parents at all affected by how intelligent they are?" (p. 203). The authors review literature that parenting styles differ with social class. Middle-class parents reason with their children and appeal to abstract principles, whereas working-class parents use physical punishment. Middle-class parents encourage children to ask questions and give explanations; working-class parents expect compliance without question. In terms of abuse and neglect, both are concentrated in lower socioeconomic classes. Herrnstein and Murray suggest that these studies have neglected the possibility that parental IQ rather than status plays a role in parenting style. To support this thesis, analysis of the NLSY showed that the probability of a white mother having a low-birth-weight baby decreased as the mother's IQ went from low to high. In contrast, variation in the mother's socioeconomic background did not affect the probability of a low-birth-weight baby. Poverty did not play a statistically significant role. Another analysis showed that the probability that a child will live in poverty during the first three years of life varied with both mother's IQ and socioeconomic background. Preexisting poverty accounted for the relationship with mother's socioeconomic background, but not cognitive ability. Both mother's IQ and socioeconomic background were found to have a moderate relationship with developmental problems in children. Finally, Herrnstein and Murray showed that the probability of having a child in the bottom decile of IQ decreased markedly as the mother's IQ went from low to high and moderately as mother's SES background went from low to high. The conclusion offered by the authors is that "people with low cognitive ability tend to be worse parents" (p. 232).

In "Crime," Herrnstein and Murray open the chapter with a discussion of the dichotomy of theories of criminal behavior: psychological and sociological. Sociological theories were prevalent during the 1950s through the 1970s. Psychological theories were prevalent early in the century. Herrnstein and Murray argue that crime is probably a function of both; yet public perceptions tend to be at the sociological pole. Previous research has suggested that criminals have a lower than average IQ. The authors explored the role that cognitive ability may play in creating criminals using the NLSY. Consistent with previous research, the mean IQ of white males sentenced to a correctional facility was

lower than the mean IQ of white males who either had no involvement with the law or were stopped by the police. Other analyses showed that variance in IQ was more strongly related to the probability of being in the top decile of self-reported crime and ever being interviewed in a correctional facility than was parental SES. A low IQ was a significant risk factor. In conclusion, Herrnstein and Murray offer the caution that "despite the relationship of low IQ to criminality, the great majority of people with low cognitive ability are law abiding" (p. 251) and the conclusion that "in trying to understand how to deal with the crime problem, much of the attention now given to problems of poverty and unemployment should be shifted to another question altogether: coping with cognitive disadvantage" (p. 251).

In "Civility and Citizenship," Herrnstein and Murray argue that cognitive ability makes a contribution to the capacity for civility and citizenship (p. 253). In particular, behavior such as voting indicates an involvement in the welfare of the community. Voting varies with socioeconomic class: "college graduates vote more than high school graduates; white-collar workers vote more than blue-collar workers; and the rich vote more than the poor" (p. 258). Indirect evidence suggests that there is an effect of cognitive ability. Analysis of the middle-class values index in the NLSY was conducted as a proxy for involvement in civility. The analysis showed that as both IQ and parental SES went from low to high, the probability of a score of "Yes" on the middle-class values (MCV) index increased. In conclusion, Herrnstein and Murray suggest that "a smarter population is more likely to be, and more capable of being made into, a civil citizenry" (p. 266).

Part III: The National Context

In this section, Herrnstein and Murray describe "ethnic differences in cognitive ability and social behavior, the effects of fertility patterns on the distribution of intelligence, and the overall relationship of low cognitive ability to what has become known as the underclass" (p. 267). Given the controversial nature of these topics, the authors request that this section be read carefully.

In Chapter 13, "Ethnic Differences in Cognitive Ability," the authors convey the message that "ethnic differences in cognitive ability are real and have consequences" (p. 269). The purposes of the authors in this chapter are twofold:

> our primary purpose is to lay out a set of statements, as precise as the state of knowledge permits, about what is currently known about the size, nature, validity, and persistence of ethnic differences on measures of cognitive ability. A secondary purpose is to try to induce clarity in ways of thinking about ethnic differences, for discussions about such differences tend to run away with themselves, blending issues of fact, theory, ethics, and public policy that need to be separated. (p. 270)

Herrnstein and Murray begin by reviewing evidence about the size of ethnic differences in cognitive ability. In general they conclude that Asians and Asian Americans have a higher average IQ than white Americans, and white Americans have a higher mean IQ than black Americans. The difference in IQ between white and black Americans is given as approximately 1.08 standard deviations. In the NLSY, "a person with the black mean was at the 11th

percentile of the white distribution, and a person with the white mean was at the 91st percentile of the black distribution" (p. 278).

The most commonly held view of this black-white difference is that it represents bias in the test. Herrnstein and Murray suggest that the difference is not a function of either predictive or cultural bias. First, "overwhelmingly, the evidence is that the major standardized tests used to help make school and job decisions do not under predict job performance" (p. 281). Second, "the B/W difference is wider on items that appear to be culturally neutral than on items that appear to be culturally biased" (p. 282). Nor does a lack of motivation to try appear to account for the difference. Herrnstein and Murray discuss the possibility of a uniform background bias, "in other words, the tests may be biased against disadvantaged groups, but the traces of bias are invisible because the bias permeates all areas of the group's performance" (p. 285). Finally, the authors consider the role of SES in generating the black/white difference in cognitive ability. First, when the two groups are matched on SES, the difference in IQ scores between the groups shrinks; however, Herrnstein and Murray point out that because SES is a result of cognitive ability, controlling for SES in this manner is guaranteed to reduce IQ differences. Second, although black IQ scores increase with SES, the magnitude of the black/white IQ difference does not decrease.

The answer to the question of whether the black/white difference is diminishing is "Yes." A review of the white/black difference in standard deviations from longitudinal data of the National Assessment of Educational Progress showed a decrease in the difference on tests of science, math, and reading across 9-, 13-, and 17-year-olds. Herrnstein and Murray speculate that the convergence may reflect rising investments in education and improvements in nutrition, shelter, and health care that disproportionately benefit cognitive development at the low end of SES (pp. 292-293). Despite the narrowing of the difference, whether convergence will continue is a source of speculation.

Next Herrnstein and Murray discuss the nature of the difference in cognitive ability between the races. The reasons given by the authors for tackling this controversial topic are to confront the assumption of genetic cognitive equality among the races that has practical consequences for society and to bring to the level of public discussion a topic that is taboo. "Taboos breed not only ignorance but misinformation" (p. 297). Finally, the authors acknowledge that

> the evidence about ethnic differences can be misused, as many people say to us. Some readers may feel that this danger places a moral prohibition against examining the evidence for genetic factors in public. We disagree, in part because we see even greater dangers in the current gulf between public pronouncements and private beliefs. (pp. 297-298)

The first point that Herrnstein and Murray emphasize is "that a trait is genetically transmitted in individuals does not mean that group differences in that trait are also genetic in origin" (p. 298). From this point, the authors estimate that if the group differences are a function of environmental differences, then "the mean environment of whites is 1.58 standard deviations better than the mean environment of blacks and .32 standard deviation worse than the mean environment for East Asians" (p. 298). In the opinion of the authors,

environmental differences of this magnitude are implausible yet not impossible; however, for those who hold that the IQ difference is due to environmental differences, "explanations have to be formulated rather than simply assumed" (p. 299).

Herrnstein and Murray provide a case for thinking that genetics may be involved in the group differences. Profile differences between the groups are cited as evidence for a genetic factor. East Asians typically score at the same level or slightly lower than whites on verbal IQ subtests, but score much higher on subtests of visuospatial IQ. Blacks typically score higher than whites on subtests involving arithmetic and immediate memory whereas whites typically score higher than blacks on subtests of spatial-perceptual ability. Furthermore, observations consistent with Spearman's hypothesis that the greater the degree to which a test measures *g*, the greater the black-white difference in test scores have been shown. The conceptual linkage to a genetic explanation comes from the statement that *g* has a high degree of heritability. Arguments against a genetic factor in group differences are reviewed. Briefly, these include uniform cultural bias, the Flynn effect (i.e., test scores tend to rise in the intervals between standardization), and transracial adoption studies.

In a final section, Herrnstein and Murray summarize the preceding discussion with the following:

> If the reader is now convinced that either the genetic or environmental explanation has won out to the exclusion of the other, we have not done a sufficiently good job of presenting one side or the other. It seems highly likely to us that both genes and environment have something to do with racial differences. What might the mix be? We are absolutely agnostic on that issue; as far as we can determine, the evidence does not yet justify an estimate.
>
> We are not so naive to think that making such statements will do much good. People find it next to impossible to treat ethnic differences with detachment. That there are understandable reasons for this only increases the need for thinking clearly and with precision about what is and is not important. In particular, we have found that the genetic aspect of ethnic differences has assumed an overwhelming importance. One symptom of this is that while this book was in preparation and regardless of how we described it to anyone who asked, it was assumed that we are faced with two alternatives: either (1) the cognitive difference between blacks and whites is genetic, which entails unspoken but dreadful consequences, or (2) the cognitive difference between blacks and whites is environmental, fuzzily equated with some sort of cultural bias in IQ tests, and the difference is therefore temporary and unimportant. (pp. 311-312)

Further on, Herrnstein and Murray discuss the hypothetical consequences of learning that ethnic differences in intelligence are genetic in origin.

> If it were known that the B/W difference is genetic, would I treat individual blacks differently from the way I would treat them if the differences were environmental? Probably, human nature being what it is, some people would interpret the news as a license for treating all whites as intellectually superior to all blacks. But we hope that putting this possibility down in words makes it obvious how illogical—besides utterly unfounded—such reactions would be. (p. 312)

Second, the authors confront the common assumption that "if the differences are genetic, aren't they harder to change than if they are environmental?" (p. 313). The error here, according to the authors, is the underlying assumption that environmentally induced deficits are less hard-wired and less real than genetically induced deficits (p. 313). In sum, Herrnstein and Murray state that knowledge that the differences are wholly genetic as opposed to wholly environmental does not justify treating another individual differently.

In Chapter 14, "Ethnic Inequalities in Relation to IQ," Herrnstein and Murray offer a view of ethnic differences in the social behaviors previously discussed in Part II when cognitive ability is taken into account. Specifically, the probabilities of various outcomes are compared for whites, blacks, and Latinos before and after controlling for IQ. After controlling for IQ, ethnic differences in higher education, occupations, and wages diminish. Differences in the probability of living in poverty and of being unemployed also shrink when controlling for IQ. Ethnic differences in marriage rates do not change. For illegitimacy, the difference between whites and Latinos decreases; however, the difference between blacks and whites remains large. Disparities between the different ethnic groups in the probability that a woman has been on welfare also diminish when controlling for IQ. The disparity between blacks and whites in the probability of giving birth to a low-birth-weight baby also decreases. Ethnic differences in a child living in poverty for the first three years of life also diminish. Differences in the probability of incarceration between the different groups also decline. This chapter portrays the probability of various outcomes for members of different ethnic groups of similar cognitive ability. From this perspective, many ethnic differences in social and economic indicators diminish while others remain.

In Chapter 15, "The Demography of Intelligence," Herrnstein and Murray explore the downward pressure exerted on the national distribution of cognitive ability by differences in birth rates across the range of cognitive ability.

> In particular, if women with low scores reproduce more rapidly than women with high scores, the distribution of scores will, other things equal, decline, no matter whether the women with the low scores came by them through nature or nurture. (p. 342)

Prior to modernization, higher status females had a reproductive advantage over lower status females. More children of higher status females survived than did children of lower status females. During this period large families were the norm. With modernization, the birth rates of privileged females declined disproportionately as these women put off marriage and reproduction to take advantage of educational and career opportunities (p. 344). Birth rates among less educated and unprivileged females remained high due to the intrinsically rewarding nature of motherhood. Consequently, "if reproductive rates are correlated with income and educational levels, which are themselves correlated with intelligence, people with lower intelligence would presumably be outreproducing people with higher intelligence and thereby producing a dysgenic effect" (p. 345).

"Foretelling the future about fertility is a hazardous business, and foretelling it in terms of IQ points per generation is more hazardous still" (p. 348). Although Herrnstein and Murray decline to forecast the future, they do com-

ment on the current status of the number of children currently born to women at various IQ levels, the age at which they bear children, and the cognitive ability of immigrants. To illustrate the relation between IQ and fertility, Herrnstein and Murray show that in 1992, the mean number of children ever born to women ages 35 to 44 decreased as level of educational attainment increased. Furthermore, 71% more births occurred among high school dropouts than among college graduates (p. 349). The dysgenic effect is also influenced by the low IQ group having children at younger ages than the higher IQ group. Consequently, the low IQ group will produce more generations per unit time than will the high IQ group. In the NLSY, the mean age at first birth for mothers at the lowest level of cognitive ability was seven years younger than for mothers at the highest level.

Second, the mean ages at which women from different ethnic groups have children also differ. From the NLSY, the mean ages for whites, Latinos, and blacks were 24.3, 23.2, and 22.3 years respectively. Herrnstein and Murray predict that if these age differentials persist, they will increase the disparity in the cognitive ability of successive generations (p. 355). Finally, the authors suggest that immigration that constituted 29% of the population growth in the 1980s also contributes to the downward pressure on the national distribution of intelligence. Herrnstein and Murray estimate that the mean IQ of immigrants from the 1960s through the 1980s was less than the mean IQ of the native-born American population.

> Putting the pieces together—higher fertility and a faster generational cycle among the less intelligent and an immigrant population that is probably somewhat below the native-born average—the case is strong that something worth worrying about is happening to the cognitive capital of the country. (p. 364)

In Chapter 16, "Social Behavior and the Prevalence of Low Cognitive Ability," Herrnstein and Murray illustrate the prevalence of low cognitive ability among people who suffer most from the social problems outlined in Part II. Using the NLSY the percentages of individuals in poverty, permanent high school dropouts, men who worked 52 weeks in 1989, able-bodied men who did not work, men ever interviewed in jail, women ever on welfare, women on welfare for five or more years, children born out of wedlock, mothers with low-birth-weight babies, children living in poverty, children in the bottom decile of IQ, and percent of people scoring "Yes" on the MCV index were plotted for each decile of IQ with the cumulative percentage function also drawn. Most of the cumulative percentage functions are negatively accelerated monotonic functions. This means that the largest percentages were in the lowest decile of IQ and that the percentages decreased with increasing deciles.

Part IV: Living Together

In this final section of the book, Herrnstein and Murray attempt to outline the relevance of cognitive ability to understanding some major domestic issues in America today. In particular, the implications of cognitive ability for education and affirmative action are discussed.

In Chapter 17, "Raising Cognitive Ability," Herrnstein and Murray review the record of attempts to raise cognitive ability through environmental improvements. "Taken together, the story of attempts to raise intelligence is one

of high hopes, flamboyant claims, and disappointing results" (p. 389). Attempts to raise cognitive ability through nutrition have shown that vitamin and mineral supplements may produce increases in scores on nonverbal subtests. However, as Herrnstein and Murray point out, this finding should be regarded with caution as many variations of successful studies failed to replicate the effect on IQ. Most attempts to raise cognitive ability have focused on improvements in education. Of such attempts, Herrnstein and Murray conclude that "although they make some differences in IQ, the size of the effect is small" (p. 394). Studies of natural variation in quality of education support this conclusion. The Coleman report concluded that "variations in teacher credentials, per pupil expenditures, and other objective factors in public schools do not account for much of the variation in cognitive abilities of American school children" (p. 395). Compensatory education programs designed to improve the cognitive functioning of disadvantaged students generally produced no effect of narrowing the gap of cognitive ability. According to Herrnstein and Murray, evidence of interventions that produce measurable improvements in IQ comes from two sources. A controlled experiment in Venezuela in which additional lessons were given to students in an experimental group over a period of a year increased their IQ scores between 1.6 and 6.5 points compared with a control group. Second, studies of the effects of commercial coaching for the SAT show a relationship between hours of study and increases in the SAT Math and Verbal scores. In sum, Herrnstein and Murray offer that "as of now, the goal of raising intelligence among school-age children more than modestly, and doing so consistently remains out of reach" (p. 402).

Preschool programs designed to raise cognitive ability such as Head Start produce short-term increases in cognitive ability that fade out over time. A naturalistic intervention that has shown promise for raising cognitive ability is "adoption out of a bad environment into a good one" (p. 410). Herrnstein and Murray review two studies that suggest that adoption from low SES to high SES homes produces a 12-point gain in IQ. In conclusion, the authors advocate good nutrition for all children, that Head Start programs be recognized for "rescuing small children from unsuitable, joyless, and dangerous environments" (p. 415), and that "the school is not a promising place to try to raise intelligence or to reduce intellectual differences" (p. 415). Advances in the endeavor to raise intelligence must await new insights about the development of cognitive ability.

In Chapter 18, "The Leveling of American Education," Herrnstein and Murray propose that the problem with American education is reflected in the declining SAT scores among the most gifted students. The educational system has been "dumbed down" to meet the needs of average and below average students, but as a result the potential of the most gifted students remains undeveloped. Yet society depends on the skills of those in the top level of cognitive ability to "create its jobs, expand its technologies, cure its sick, teach in its universities, administer its cultural and political and legal institutions" (p. 418).

Herrnstein and Murray begin by showing that the common perception that the academic performance of the *average* student is much worse today is not substantiated by longitudinal measures of academic achievement. However,

for the pool of youths in the top 10-20% of cognitive ability the opposite is true. SAT scores declined markedly from 1963 to 1980. The explanation most commonly offered for this decline is an expansion of the pool of students taking the test to include students who previously did not consider going to college. According to Herrnstein and Murray, this commonly held explanation is mistaken. As evidence the authors point out that "throughout most of the white SAT score decline, the white SAT pool was shrinking, not expanding" (p. 426). Instead, the authors suggest that the decline reflects "a downward trend of the educational skills of America's most academically promising youngsters toward those of the average student" (p. 427). Further analysis of the SAT scores of the most gifted youths reveals that the number of youths scoring 700 or higher on the Verbal SAT decreased by 41% from 1972 to 1993. Math scores rebounded from the decline with an increase in the proportion of 17-year-olds scoring 700 or more during the 1980s.

Herrnstein and Murray suggest that the decline in the 1960s reflects a "dumbing down" of education.

> One of the chief effects of the educational reforms of the 1960s was to dumb down elementary and secondary education as a whole, making just about everything easier for the average student and easing the demands on the gifted student. (p. 430)

Traditional criteria of rigor and excellence gave way to "the need to minimize racial differences in performance measures, and enthusiasm for fostering self-esteem independent of performance" (p. 432). Other forces that eroded verbal skills were television replacing newsprint as a source of news and information and the telephone replacing letterwriting as the major form of communication. Math scores were less susceptible to these pressures than were verbal scores. Dilution of the curriculum benefited the mediocre student, but depressed the development of the intellectual skills of the talented student.

In concluding this chapter Herrnstein and Murray suggest that "critics of American education must come to terms with the reality that in a universal education system, many students will not reach the level of education that most people view as basic" (p. 436). One tendency is to blame students who do not work hard enough for the shortcomings of the educational system. However, the authors suggest two reasons why students are less to blame for not working harder. One is that "most American parents do not want drastic increases in the academic workload" (p. 437) and the second is that "the average American student has little incentive to work harder than he already does in high school" (p. 437). With respect to government interventions, Herrnstein and Murray recommend that "the federal government should actively support programs that enable all parents, not just affluent ones, to choose the school that their children attend" (p. 440), that the government should establish a federal scholarship program for students earning the top scores on standardized tests of academic achievement, and the government should "reallocate some portion of existing elementary and secondary school federal aid away from programs for the disadvantaged to programs for the gifted" (pp. 441-442). Finally, Herrnstein and Murray recommend the resurrection of the classical idea of the educated person.

> To be an educated person must mean to have mastered a core of history, literature, arts, ethics, and the sciences and, in the process of learning those disciplines, to have been trained to weigh, analyze, and evaluate according to exacting standards. (p. 444)

In the next chapter, "Affirmative Action in Higher Education," Herrnstein and Murray provide an overview of affirmative action in practice and argue that "current practice is out of keeping with the rationale for affirmative action" (p. 451). In general, most people agree that affirmative action in education in needed, but differ when asked to say what they mean by it. The authors offer the following definition: "perfectly practiced, affirmative action means assigning a premium, an edge, to group membership in addition to the individual measures before making a final assessment that chooses some people over others" (p. 450).

Herrnstein and Murray examine the magnitude of this edge in SAT scores for blacks, whites, and Asians at 16 of 20 top-rated universities. The median difference between the white and black means was 180 SAT points whereas the median difference between the Asian and white means was 30 SAT points. On the LSAT, the differences between the white mean and the means for blacks, Latinos, and Asian were –1.49, –1.01, and –.32 standard deviations respectively. Differences of similar magnitude were found for MCAT and GRE scores.

To answer the question of whether these differences that reflect affirmative action in practice are good or bad, Herrnstein and Murray examine the logic of college admissions. "College admission is not, has never been, nor is there reason to think that it should be, a competition based purely on academic merit" (p. 459). Rationales for affirmative action in the admissions process fall into the categories of institutional benefit, social utility, and just deserts. Institutional benefit refers to the rationale to admit students from racial and ethnic minorities to enrich the campus by adding to its diversity (p. 459). Social utility refers to the rationale to admit minority youths to increase minority representation at higher socioeconomic and professional levels, so that in the future it will be easier for minorities to pursue such professions. Just deserts refers to taking into account how well an applicant has done given the environment that it was accomplished in.

> The applicant who overcame poverty, cultural disadvantages, an unsettled home life, a prolonged illness, or a chronic disability to do as well as he did in high school will get a tip from most admissions committees, even if he is not doing as well academically as the applicants usually accepted. (p. 461)

To examine how these rationales would operate in theory in a selection process, Herrnstein and Murray set up a table for deciding between a white candidate and a minority candidate with SES for each candidate either low or high. For the decision between a high SES white and a low SES minority, the rationales favor assigning a large preference for the minority candidate. For the case of a high SES white and a high SES minority candidate, the just deserts and social utility rationales are in opposition, but a small preference is assigned to the minority. For the decision between a low SES minority and a low SES white candidate, the rationales favor both candidates; therefore, little to no premium is given to the minority candidate. For the case of the high SES minority over

the low SES white candidate, social utility favors both, but just deserts favor the white candidate so a modest premium is assigned to the white candidate. Using the NLSY, the authors show that for affirmative action in practice, the differences in cognitive ability in standard deviation units between a black candidate and a white candidate for the four cells described above were +1.25, +.91, +1.17, and +.58, respectively. In each case, the analysis indicates a large edge in cognitive ability is given to the minority candidate. To the reader, Herrnstein and Murray ask in a properly run system of affirmative action, "How big an edge is appropriate?" (p. 467).

Another consideration is the cost of affirmative action. "How much harm is done to minority self-esteem, to white perceptions of minorities, and ultimately to ethnic relations by a system that puts academically less able minority students side by side with students who are more able?" (p. 470). Students commonly observe the racial mix of the student population, students who stand out because they seem out of place in college, and students who stand out because they are especially smart (p. 471). Minority students may stand out due to poor academic performance. Dropout rates from college by black students are twice the rate for whites. "Getting discouraged about one's capacity to compete in an environment may be another cost of affirmative action" (p. 473). Growing racial animosity on campuses may be yet another cost. In conclusion, Herrnstein and Murray advocate a return to the original conception of affirmative action: "to cast a wider net, to give preference to members of disadvantaged groups, whatever their skin color, when qualifications are similar" (p. 448).

In Chapter 20, "Affirmative Action in the Workplace," Herrnstein and Murray evaluate the impact of affirmative action legislation on the workplace. The authors state that job discrimination legislation was developed based on

> the assumptions that (1) tests of general cognitive ability are not a good way of picking employees, (2) the best tests are ones that measure specific job skills, (3) tests are biased against blacks and other minorities, and (4) all groups have equal distributions of cognitive ability. (p. 483)

Herrnstein and Murray state that although these assumptions were defensible back in the 1960s, today it is well established that tests of cognitive ability are related to job productivity, that these tests have more predictive power than grades, education, or a job interview; that this predictive power arises from their measure of general cognitive ability rather than specific skills, that these tests are not biased against blacks, and that different ethnic groups have different distributions of cognitive ability (pp. 483-484). The authors recommend that "the government should scrap the invalid scientific assumptions that undergird policy and express policy in terms that are empirically defensible" (p. 484).

Next Herrnstein and Murray evaluate the effects of affirmative action in the workplace without and with controlling for cognitive ability. Without controlling for IQ, there has been an increasing trend in the percentage of blacks employed in clerical, professional, and technical jobs from 1960 to 1990. No trend shows an effect of the antidiscrimination law on hiring and promotions. When the same trend lines are adjusted for the known difference in IQ between blacks and whites, the trend lines show that in both clerical and professional

and technical positions, for individuals in the same IQ range, blacks were being hired at higher rates than whites since the 1960s with both trends increasing into the 1980s. Herrnstein and Murray suggest that when adjusted for IQ, evidence for the impact of affirmative action becomes apparent. In the 1950s, blacks were systematically and unjustly excluded from these occupations. In the mid 1960s the underrepresentation of blacks in technical and professional occupations disappeared, and since the 1960s the representation of blacks in these professions has gone beyond parity (pp. 491-492).

In conclusion, Herrnstein and Murray advocate that the goal of an affirmative action policy should be equality of opportunity rather than equality of outcome. Furthermore, they state that

> if the quality of performance fairly differs among individuals, it may fairly differ among groups. If a disproportion is fair, then "correcting" it—making it proportional—may produce unfairness along with equal representation. We believe that is what happened in the case of current forms of affirmative action. (p. 501)

From this point, Herrnstein and Murray offer four policy alternatives that they think will produce fair treatment in hiring and promotions. One alternative is to create tests that meet the current requirements but still predict job performance. The problem with this alternative is that the relation of general ability to job performance will produce disparate impacts across groups. A second alternative is to allow employers to use educational credentials to narrow the pool of qualified applicants. The problem with this approach is that equivalent degrees do not mean equivalent cognitive ability due to affirmative action in universities. The third alternative is to produce norms for each group. Scores would be converted to percentiles based on the distribution for each group and employers could hire top down from each distribution. According to the authors, race norming was used in the early 1980s but was outlawed in 1986. Finally, the fourth proposal, which is favored by the authors, is based on the proposition that "if tomorrow all job discrimination regulations based on group proportions were rescinded, the United States would have a job market that is ethically fairer, more conducive to racial harmony, and economically more productive, than the one we have now" (p. 505). As with education, Herrnstein and Murray advocate that the government "get rid of preferential affirmative action and return to the original conception of casting a wider net and leaning over backward to make sure that all minority applicants have a fair shot at the job or the promotion" (p. 505).

In the next chapter, entitled "The Way We Are Headed," Herrnstein and Murray predict a pessimistic future for American society based on the following trends "an increasingly isolated cognitive elite, a merging of the cognitive elite with the affluent, a deteriorating quality of life for people at the bottom end of the cognitive ability distribution" (p. 509). Throughout this century, American society has been transformed into a society stratified by cognitive ability. The cognitive elite has grown to become a new class of individuals in high IQ occupations that has displaced previous socioeconomic elites. The authors state that "the invisible migration of the twentieth century has done much more than let the most intellectually able succeed more easily. It has also segregated them and socialized them" (p. 513). Furthermore, the very bright have become more affluent and the affluent increasingly comprise the very

bright, which has led to a blending of the interests of the affluent and the cognitive elite (p. 515). According to the authors, this will lead to what Robert Reich has referred to as a "secession of the successful" (p. 517). A scenario is presented where 10-20% of the population are wealthy enough to bypass social institutions that they do not agree with and to constitute a political bloc with considerable clout. Herrnstein and Murray suggest that this new coalition of the affluent and the intelligent will fear the growing underclass in American society.

For the underclass, the future predicted by the authors is grim.

> People in the bottom quartile of intelligence are becoming not just increasingly expendable in economic terms; they will sometime in the not-too-distant future become a net drag. In economic terms and barring a profound change in direction for our society, many people will be unable to perform that function so basic to human dignity: putting more into the world than they take out. (p. 520)

As the illegitimacy rate among whites increases, a white underclass will emerge that will be subject to much the same fear and resentment from the cognitive elite that the black underclass experiences. Meanwhile, the outmigration of the ablest blacks from the black inner city will lead to an increasing concentration of blacks with limited cognitive ability and the attendant social problems (p. 522).

Finally, Herrnstein and Murray foresee that "the cognitive elite, with its commanding position, will implement an expanded welfare state for the underclass that also keeps it out from underfoot" (p. 523). Some of the features of that welfare state will be "child care in the inner city will become primarily the responsibility of the state" (p. 523), "the homeless will vanish" (p. 523), "strict policing and custodial responses to crime will become more acceptable and widespread" (p. 524), "the underclass will become even more concentrated spatially than it is today" (p. 524), "the underclass will grow" (p. 525), "social budgets and measures for social control will become still more centralized" (p. 525), and "racism will reemerge in a new and more virulent form" (p. 525). To avoid this future, Herrnstein and Murray recommend that the issues of a society increasingly dominated by a cognitive elite and a growing underclass need to be addressed now.

The final chapter of the book is entitled "A Place for Everyone." Herrnstein and Murray state that "our central concern since we began writing this book is how people might live together in harmony despite fundamental individual differences" (p. 528). To achieve this goal, the authors advocate a return to the original conception of human equality and the pursuit of happiness. The original conception of equality of rights that arose from the philosophies of Hobbes and Locke is not seen as synonymous with the assumption that "individuals are both equal and empty, a blank slate to be written upon by the environment" (p. 529). "They are equal in rights, Locke proclaimed, though they be unequal in all other things" (p. 530). Furthermore, the authors argue that this original conception of equal rights was what the Founders of America espoused. "The Founders saw that making a stable and just government was difficult precisely because men were unequal in every respect except their right to advance their own interests" (p. 531). In contrast, contemporary political theory works with a conception of equality that seeks to suppress social and

economic inequalities that result when people are free to behave differently (p. 532).

Herrnstein and Murray suggest that for society to come to terms with the realities that people differ in intelligence and that intelligence bears on how well people do in life, it should operate in such a manner so as to allow individuals throughout the range of cognitive ability to find valued places for themselves (p. 535). In the past,

> when the responsibilities of marriage and parenthood were clear and uncompromising and when the stuff of community life had to be carried out by the neighborhood or it wouldn't get done, society was full of accessible valued places for people of a broad range of abilities. (pp. 537-538)

In this traditional context, it was easier for an individual with low cognitive ability to find a valued place. In the contemporary world, the increased costs of the valued roles of spouse, parent, and neighbor as well as the stripping of traditional functions from the neighborhood and the community has made it increasingly more difficult for individuals of modest cognitive ability to find valued places in society (p. 538). From this observation, Herrnstein and Murray propose that "a wide range of social functions should be restored to the neighborhood when possible and otherwise to the municipality" (p. 540) in order to increase the valued places that people can fill.

The second set of policy prescriptions advocated by Herrnstein and Murray involve a simplification of rules generated by the cognitive elite that make life more difficult for everyone else (p. 541). "As the cognitive elite busily goes about making the world a better place, it is not so important to them that they are complicating ordinary lives" (p. 541). Such complexity serves as a barrier to people who are not cognitively equipped to struggle through the bureaucracy and, according to Herrnstein and Murray, should be removed. Another prescription concerns making the justice system simpler so that the rules about crime and the consequences for crime are simpler.

> The number of acts defined as crimes has multiplied, so that many things that are crimes are not nearly as obviously "wrong" as something like robbery or assault. The link between moral transgression and committing crime is made harder to understand. (p. 543)

Such simplification will make living a moral life simpler for persons of lower cognitive ability. For marriage, Herrnstein and Murray argue that the sexual revolution and the state have made it "much more difficult for a person of low cognitive ability to figure out why marriage is a good thing, and, once in a marriage, more difficult to figure out why one should stick with it through bad times" (p. 544). As a policy prescription, Herrnstein and Murray advocate returning marriage to its historic legal status to restore the rewards of marriage by validating the rewards that marriage naturally carries with it (p. 546).

Herrnstein and Murray conclude the book with the following words:

> Cognitive partitioning will continue. It cannot be stopped, because the forces driving it cannot be stopped. But America can choose to preserve a society in which every citizen has access to the central satisfaction of life. Its people can, through an interweaving of choice and responsibility, create valued places for themselves in their worlds. They can live in communities—urban or rural—

where being a good parent, a good neighbor, and a good friend will give their lives purpose and meaning. They can weave the most crucial safety nets together, so that their mistakes and misfortunes are mitigated and withstood with a little help from their friends.

All of these good things are available now to those who are smart enough or rich enough—if they can exploit the complex rules to their advantage, buy their way out of social institutions that no longer function, and have access to the rich human interconnections that are growing, not diminishing, for the cognitively fortunate. We are calling upon our readers, so heavily concentrated among those who fit that description, to recognize the ways in which public policy has come to deny those good things to those who are not smart enough and rich enough.

At the heart of our thought is the quest for human dignity. The central measure of success for this government, as for any other, is to permit people to live lives of dignity—not to give them dignity, for that is not in any government's power, but to make it accessible to all. That is one way of thinking about what the Founders had in mind when they proclaimed, as a truth self-evident, that all men are created equal. That is what we have in mind when we talk about valued places for everyone. (p. 551)

Reference

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 257-264*

Douglas Wahlsten
University of Alberta

Increasing the Raw Intelligence of a Nation is Constrained by Ignorance, Not its Citizens' Genes

In The Bell Curve, *Herrnstein and Murray claim that a high value for heritability of intelligence limits or constrains the extent to which intelligence can be increased by changing the environment. This article argues that the concept of heritability is based on unsupportable assumptions and that its numerical value places no constraint on the consequences of an improved environment. On the contrary, a very small change in environment, such as a dietary supplement, can lead to a major change in mental development, provided the change is appropriate to the specific kind of deficit that in the past has impaired development. The results of adoption studies and the intergenerational cohort effect also reveal that intelligence can be increased substantially without the need for heroic intervention.*

The Bell Curve by Herrnstein and Murray (1994) addresses what they and many politicians perceive as a crisis in the United States brought about by wrong-headed government policies, policies they claim have resulted in "disaster," and they urge that government leaders should "try living with inequality" rather than striving to eradicate it. They believe that many people lack sufficient intelligence to be successful in American society, and they argue that an important part of their inability to compete successfully arises from inadequate genes.

Even before it appeared in retail stores, the book was prominently publicized in the mass media as offering important insights about genetic causes of low intelligence and poverty. The October 16, 1994 issue of the *New York Times Book Review* featured *The Bell Curve* with a cover picture of a DNA double helix and the question "How much of us is in the genes?" Inside was a sympathetic review (Browne, 1994) emphasizing "ineradicable cognitive disability created by genetic bad luck" (p. 3). Although data in the book are almost exclusively concerned with the United States and the obsession with race is a peculiarly American trait, *The Bell Curve* was given major, albeit more critical, attention in national Canadian media (Bruning, 1994; Campbell, 1994). After a remarkably short delay, multiauthored, book-length discussions of *The Bell Curve* went on sale (Fraser, 1995; Jacoby & Glauberman, 1995). Herrnstein and Murray wrote a deliberately provocative book and in this respect they succeeded marvelously. Vast numbers of intellectuals in Canada and the US have set aside their work for a while and occupied their minds with genes and psychological testing. Although the book was not written to advance scientific knowledge and will not have much appeal for those in the biological sciences

Douglas Wahlsten is a professor of psychology, having moved from the University of Waterloo in 1989. Since 1969 he has studied genetic aspects of learning and brain structure with the support of NSERC and MRC grants.

in particular, it has become necessary for many of us to examine it in some detail because of persistent questions about it from colleagues and journalists.

As a specialist in behavioral and neural genetics, I focused on those chapters involving claims about heredity and inequality. Herrnstein specialized in the study of learning in rats and pigeons, whereas Murray is a political scientist. Herrnstein and Murray are obviously not geneticists: they are writing for an audience that knows little about genetics, and their readers will not gain a better understanding of genetic principles from reading their book. On the contrary, diligent readers who try to follow their reasoning will probably become confused and misled about the role of genes and the root causes of social problems.

Herrnstein and Murray claim that statistical methods can reveal with reasonable accuracy the percentage of individual variation in intelligence that is caused by genetic differences among people, and they conclude that the "most unambiguous direct estimate" indicates this percentage is about 60-70%. They do not mention that this methodology presumes the effects of genes and environment occur separately during development and combine by simple arithmetical summation, or that this presumption has been rejected as biologically unrealistic by many geneticists (Gottlieb, 1992; Lewontin, 1974; McGuire & Hirsch, 1977; Wahlsten, 1990, 1994). They claim that high "heritability" of IQ means that improving the environment of a poor child a modest amount will be ineffective because "such changes are limited in their potential consequences when heritability so constrains the limits of environmental effects" (p. 109).

The Bell Curve is simply wrong on this point. A heritability estimate does not in any way constrain the effects of a moderately changed environment. Small treatments tend to have small effects unless the treatments directly and precisely ameliorate a specific difficulty that impairs development. If such a specific difficulty can be identified, a very small change in the environment can lead to a dramatic improvement. For example, during the first few decades of the 20th century, pellagra was quite common among the working poor of the southern US. The eugenicist Davenport claimed the slow learning and health problems of pellagrins resulted from an infection combined with bad genes, while the experimental *proof* by the physician Goldberger that it was a vitamin deficiency disease caused by low wages leading to a poor diet was ignored (Chase, 1977). Now we know that a small daily dose of the vitamin niacin can effectively prevent pellagra, just as vitamin C prevents scurvy and low phenylalanine milk prevents symptoms of the genetic disease phenylketonuria (PKU). Sweeping statements about the ineffectiveness of environmental change denote helplessness and pessimism occasioned by ignorance rather than any inherent resistance of intelligence to modification. Each of the 50,000 or more genes in the human chromosomes functions in a highly specific way as part of the biochemical system of a cell, and genetic knowledge can help to devise effective treatments only when a specific gene that impairs development is known. Bereft of genuine genetic knowledge, the kind of pseudogenetic heritability estimates espoused by Herrnstein and Murray serve as a weapon against the poor in the propaganda arsenal of reactionary politicians.

The Bell Curve asserts confidently that "Changing cognitive ability through environmental intervention has proved to be extraordinarily difficult" (p. 314).

This is false. Available data indicate that a modest, short term improvement such as the Head Start project in the US has correspondingly small effects on mental ability test scores, whereas a large and lasting improvement produced by adoption can exert quite a large effect. Well-controlled studies done in France have found that transferring an infant from a family having low socio-economic status (SES) to a home where parents have high SES improves childhood IQ scores by 12 to 16 points or about one standard deviation (Capron & Duyme, 1991; Schiff, Duyme, Dumaret, & Tomkiewicz, 1982; see Figure 1), which is considered a large effect size in psychological research (Cohen, 1988). Adoption can entail a major improvement in a child's environment, but the adoptive home is usually not off the scale of decent environments and therefore is not expected to yield a rich harvest of superior intellects. Achieving extraordinarily high levels of performance requires exceptional effort under the tutelage of expert instructors (Ericsson, Krampe, & Tesch-Römer, 1993; Wagner & Oliver, in press). Outstanding achievement and brilliant creativity do not come "naturally" to anyone merely because of their genes.

Changing mental ability test scores a modest amount is not so difficult. In fact, routine IQ testing reveals this commonly happens without deliberate intervention to enhance intelligence per se. The extent of this phenomenon

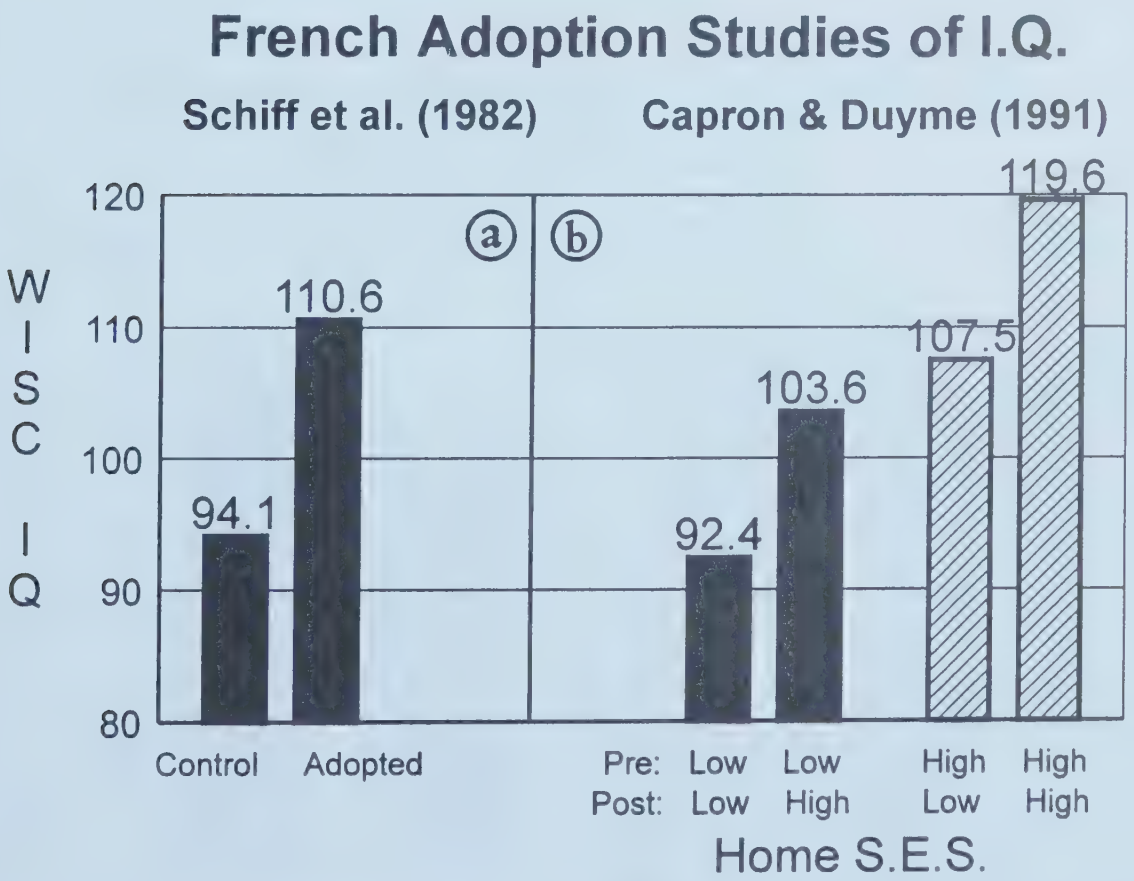


Figure 1. Mean WISC IQ score for children in two adoption studies done in France. (a) Schiff et al. (1982) compared two groups of children having the same biological mother and similar biological fathers. One group had been adopted into homes of well-educated professionals, whereas the control children had remained with the mother living in poverty. (b) Capron and Duyme (1991) tested children from four conditions categorized according to parental socioeconomic status (SES) prior to and after adoption.

tends to be obscured by the method of scoring the tests. The average IQ in a population should be about 100 and the standard deviation should be about 15, such that about 98% of people at the same age will score between 70 and 130 IQ points. One mechanism for scaling an IQ test is simple. The test is given in a particular year to a representative random sample of the population, and the numbers of items correct have mean and standard deviation of M and S, respectively. Then each raw score X is converted to a standard score Z that represents the number of standard deviations from the mean for the individual, using Z=(X–M)/S. The Z scores have a mean of 0 and standard deviation of 1. Finally, the Z scores are transformed to IQ scores with the formula IQ=100+Z(15). Thus IQ indicates relative performance on a test rather than absolute degree of intelligence. Over a period of several years it becomes necessary to restandardize the IQ test using more appropriate test items and a new sample of the population. This periodic restandardization of a test tends to keep the mean IQ close to 100, even if the underlying trait called intelligence is changing substantially in the population.

An immense body of evidence reveals that raw, unstandardized intelligence has been gradually increasing for several decades since World War II in many industrialized countries including Canada (Flynn, 1987). Two kinds of data show this trend clearly. Perhaps the most persuasive comes from the

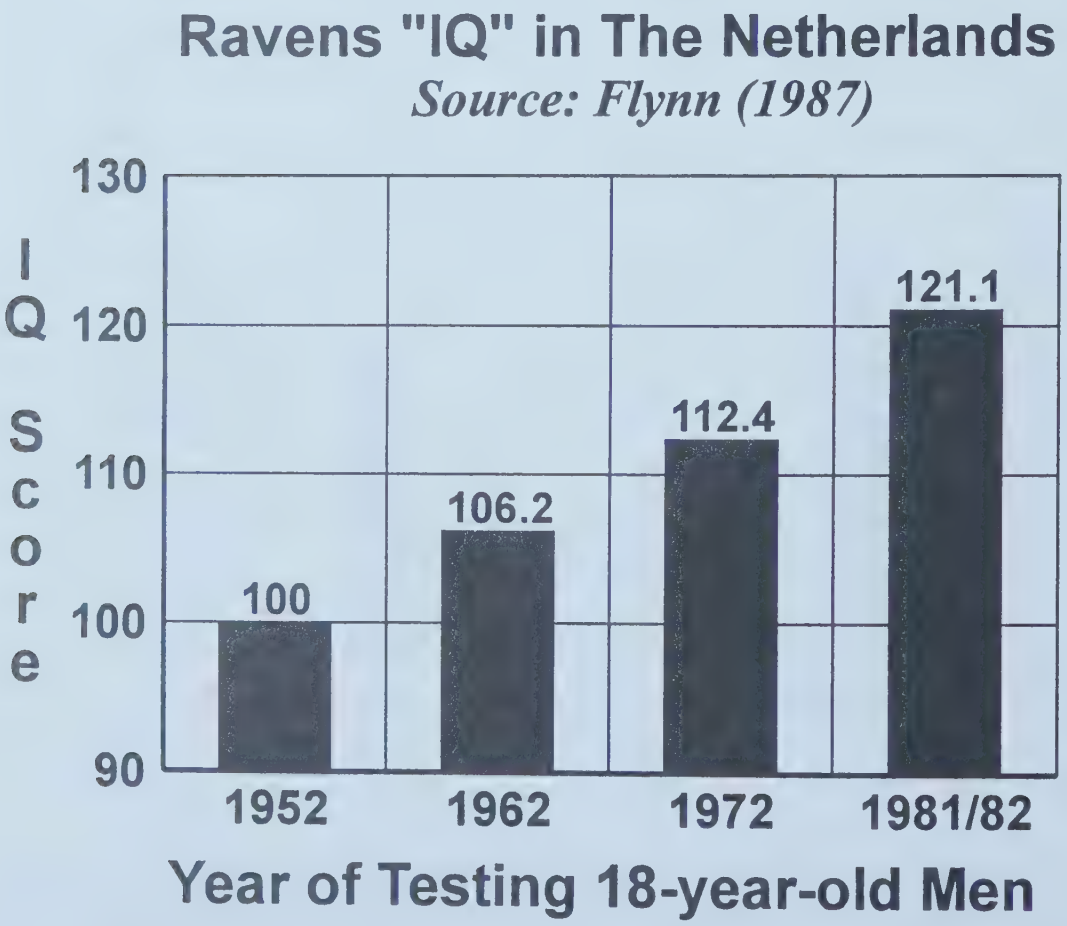


Figure 2. Mean Raven's Progressive Matrices scores converted to IQ scores for 18-year-old men in the Netherlands who were tested at the time of military induction in different years. Based on data in Table 1 of Flynn (1987).

Netherlands, an ethnically homogenous country, where almost all 18-year-old males are given the Raven's Progressive Matrices test as part of military induction. The test itself has not been modified for several decades. As shown in Figure 2, there is a very large *cohort effect* amounting to 21 IQ points increase in the population over three decades.

The other kind of evidence derives from the restandardization procedure when people given the new version of a test then take the old version so that validity can be assessed. For example, when the Wechsler Intelligence Scale for Children (WISC) was revised in 1972, the sample of children scored 7 points higher on the previous version of the WISC that had been standardized in 1947. Combining these kinds of data for several IQ tests, a one-standard-deviation increase in mean intelligence in the US is apparent over several decades (Figure 3). The cohort effect is gradual and almost linear since World War II, but in terms of a population-wide change in intelligence manifested in one generation of Americans, it is a large effect. It is especially thought-provoking that the size of the cohort effect is not much different than the widely publicized black-white IQ difference in the US. That is, more recently born children exceed the raw intelligence of their own parents at a comparable age by almost the same average amount as Americans of European ancestry exceed Americans of African ancestry, especially on more recent tests of mental ability.

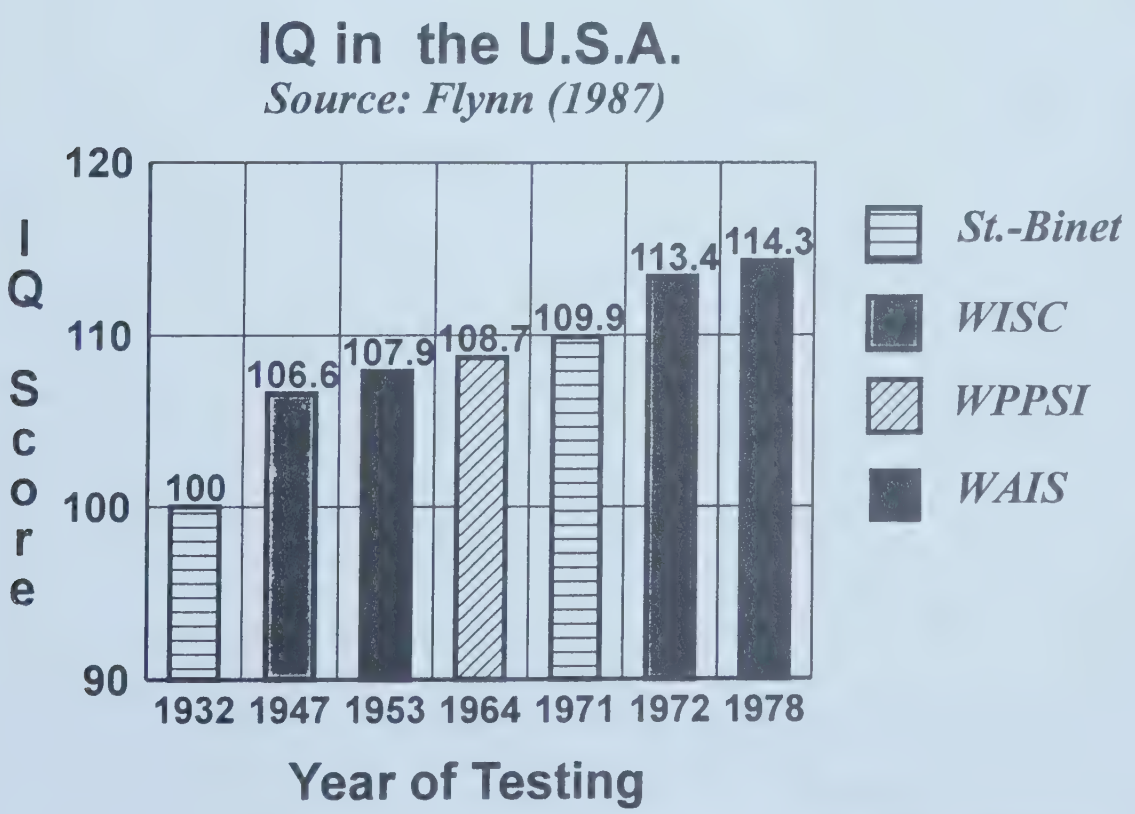


Figure 3. Mean IQ scores of standardization samples for four IQ tests given to Americans of European ancestry in various years. In each case two or more of the tests were taken by the same people in the same year, and averages were expressed in terms of the base score of 100 on the Stanford-Binet in 1932. Abbreviations: St.-Binet, Stanford-Binet; WISC, Wechsler Intelligence Scale for Children; WAIS, Wechsler Adult Intelligence Scale; WPPSI, Wechsler Preschool and Primary Scale of Intelligence. Based on data in Table 7 of Flynn (1987).

Herrnstein and Murray mention the cohort effect in their discussion of group differences, but they lightly dismiss it as a mere improvement in "test taking" skills or betterment of the living conditions of the disadvantaged. They muse that "one does not get the impression that the top of the IQ distribution is filled with more subtle, insightful, or powerful intellects than it was in our grandparents' day" (p. 308). Thus, faced with weighty evidence against their thesis, they are willing to dismiss the cohort effect with counter-arguments that negate some of their own fundamental claims. If the cohort effect represents improvement by only the bottom half of the "bell curve" rather than the entire population, then their earlier claim that increasing IQ is extraordinarily difficult loses credibility. After all, if the population mean increases by 15 points but the top part of the distribution does not increase, the lower scores must have increased by much more than 15 points. Herrnstein and Murray maintain assiduously in the first half of their book that IQ tests as we know them are very good measures of general intelligence. Yet the cohort effect causes them to revert to subjective impressions about their grandparents who must have been children when the IQ test was still embryonic in the mind of Alfred Binet. Actual experience in the US with the earliest administrations of IQ tests revealed that many men at the apex of American society were none too heavy under the helmet. The December 29, 1915 issue of the *Chicago Herald* trumpeted to its public: "Hear how Binet-Simon method classed mayor and other officials as morons" (Chase, 1977, p. 241). As for outstanding intellects, without doubt they are products of their times and countries, but their achievements do not provide a valid measure of the intelligence of their lesser countrymen, who all too often failed to recognize genius in their midst, partly because of the prevailing political and social definition of genius (Weisberg, 1986). Formal IQ tests were intended to supplant subjective impression and common prejudice with carefully constructed instruments administered in a controlled conditions. For Herrnstein and Murray to tiptoe around the cohort effect by suggesting that IQ tests do not *really* measure *genuine* intelligence but something more superficial and transitory is a negation of a fundamental part of the thesis of *The Bell Curve*.

The cohort effect poses an even greater challenge to the raison d'être of *The Bell Curve*. Herrnstein and Murray raise the alarm about several worrisome social trends in the US and argue that inadequate intelligence is the root cause of most social problems. They present striking graphs of social statistics over several decades that reveal a dramatic deterioration in American society, especially from 1960 to 1990. Over this period, we are told that the marriage rate has declined while the divorce rate has increased from 7% to 20% and "illegitimate" births have increased from 5% to 30%, welfare caseloads have risen from 1.5% to 7%, and the rate of violent crimes is now five times higher than three decades ago. Nevertheless, the raw intelligence of American youth has apparently increased a substantial amount over this same interval. A national decline in intelligence could not possibly be the basis for these negative social trends.

The primary evidence Herrnstein and Murray offer for the important influence of individual intelligence in American society is a series of positive correlations between IQ and variables such as success in school, work, and social life. Correlations are notoriously poor guides to the direction of causal

influences even when multiple regression methods are used, and this kind of information cannot distinguish between socioeconomic causes of low or high intelligence and consequences that flow from differences in intelligence at any one time. Comparisons of groups of people living in changed environments, on the other hand, can reveal the direction of causation. Adoption from a low SES home into a high SES home is clearly a change in environment that precedes and causes the change in childhood IQ, presuming the theorist will admit that even the brightest infants lack the power to choose their parents. Likewise, the cohort effect implicates nationwide environmental change as the cause of enhanced childhood intelligence. This enhanced intelligence may then consolidate and build on past achievements as the youth mature and become productive, influential members of society.

In my opinion, *The Bell Curve* from beginning to end suffers from a lack of intellectual rigor and a rather cavalier use of data mustered from here and there to bolster an obvious political agenda. The authors are worried about the growing gap between the rich and the poor in the US and the apparent disintegration of American society, and they offer some suggestions for policies that make sense from the perspective of the psychology of animal learning, Herrnstein's specialty. When they invoke genetic explanations for class and racial differences in educational and occupational achievement, however, they enter a realm where their incompetence is painfully evident (Kamin, 1995). Herrnstein, Murray, and their publicists still do not understand that genetic phenomena cannot be the root causes of short-term social trends. These socioeconomic trends are rooted in the political and economic system that prevails in the US. Moral decay does not arise from a sudden, inexplicable epidemic of genetic mutations. The obsession of so many social scientists and journalists with genetic victim blaming serves to divert attention from the real causes and cures of social ills. *The Bell Curve* itself and the attention lavished on it are symptoms of social malaise in the US.

As a final example of this malaise manifest in academic circles, consider the attempt by Herrnstein and Murray to deny the accusations of fraud against British psychologist Cyril Burt, who is widely recognized as the author of fictitious data and articles purporting to show high heritability of IQ (Hearnshaw, 1979). In a box on page 12, Herrnstein and Murray cite books by Joynson and Fletcher claiming to show that accusations against Burt are groundless. What Herrnstein and Murray fail to mention is an authoritative, peer reviewed critique by Samelson (1992) of these books that finds the accusations against Burt well substantiated and the work of his defenders shoddy and one-sided. Samelson concludes:

> What does this whole affair tell us about so-called science and scientists, insiders and outsiders, power structures and establishment climates in a profession, about "experts" whose beliefs flipflop from one side to the other, about presumably responsible editors, and finally about "quality control"? Beyond some pious words about them, we do not appear to have made much progress on these issues. (p. 231)

In this respect *The Bell Curve* is a step backward rather than a benchmark of progress in the nature-nurture debate.

References

Browne, M.W. (1994, October 16). What is intelligence, and who has it? *New York Times Book Review*, pp. 3, 41, 45.

Bruning, F. (1994, November 21). Writing off the black race. *Maclean's*, p. 13.

Campbell, M. (1994, October 29). A furor in black and white. *The Globe and Mail*, D1-2.

Capron, C., & Duyme, M. (1991). Children's IQs and SES of biological and adodptive parents in a balanced cross-fostering study. *Cahiers de Psychologie Cognitive, 11*, 323-348.

Chase, A. (1977). *The legacy of Malthus. The social costs of the new scientific racism.* New York: Knopf.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Ericsson, K.A., Krampe, R.T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363-406.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171-191.

Fraser, S. (Ed.). (1995). *The bell curve wars.* New York: Basic Books.

Gottlieb, G. (1992). *Individual development and evolution. The genesis of novel behavior.* New York: Oxford University Press.

Hearnshaw, L.S. (1979). *Cyril Burt, psychologist.* New York: Vintage Books.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American Life.* New York: Free Press.

Jacoby, R., & Glauberman, N. (Eds.). (1995). *The bell curve debate.* New York: Times Books.

Kamin, L.J. (1995). Behind the curve. *Scientific American, 272*, pp. 99-103.

Lewontin, R. (1974). The analysis of variance and the analysis of causes. *American Journal of Human Genetics, 26*, 400-411.

McGuire, T.R., & Hirsch, J. (1977). General intelligence (*g*) and heritability (H^2, h^2). In I.C. Uzgiris & F. Weizmann (Eds.), *The structuring of experience* (pp. 25-71). New York: Plenum.

Samelson, F. (1992). Rescuing the reputation of Sir Cyril Burt. *Journal of the History of Behavioral Sciences, 28*, 221-233.

Schiff, M., Duyme, M., Dumaret, A., & Tomkiewicz, S. (1982). How much could we boost scholastic achievement and IQ scores? A direct answer from a French adoption study. *Cognition, 12*, 165-196.

Wagner, R.K., & Oliver, W.L. (in press). How to get to Carnegie Hall: Implications of exceptional performance for understanding environmental influences on intelligence. In D.K. Detterman (Ed.), *Current topics in human intelligence.* Norwood, NJ: Ablex.

Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity-environment interaction. *Behavioral and Brain Sciences, 13*, 109-161.

Wahlsten, D. (1994). The intelligence of heritability. *Canadian Psychology, 35*, 244-258.

Weisberg, R.W. (1986). *Creativity: Genius and other myths.* New York: Freeman.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 265-270*

Raymond E. Fancher
York University

The Bell Curve on Separated Twins

The Bell Curve *declares that studies of separated identical twins—characterized as the "purest" of the "direct" methods available for estimating IQ heritability in the American population—point to a value between +.75 and +.80. This makes Herrnstein and Murray's working, "middle-ground" estimate of +.6 ± .2 seem conservative. In fact, however, the authors of the main study cited by Herrnstein and Murray suggest a heritability of only "two-thirds" for a population restricted to the "broad middle-class." Herrnstein and Murray completely omit to mention the numerous complicating factors in all reputable separated twin studies conducted to date that have rendered them significantly less than pure, and that have inevitably inflated the observed IQ correlations.* The Bell Curve*'s overall treatment of twin studies is inaccurate and misleading.*

This commentary focuses on a relatively small but significant detail from *The Bell Curve*: namely Herrnstein and Murray's (1994) discussion of twin studies and their implications for estimating the heritability of intelligence. Of course, the assumption that intelligence is strongly heritable within the broad American population is one of the pillars on which the main weight of *The Bell Curve*'s main argument rests. The authors propose what they call a "middle-ground" (p. 298) heritability value of .6 ± .2, a figure that they claim "does no violence to any of the competent and responsible recent estimates" (p. 108). Although lower and more approximate than the estimate of .8 ± .1 proposed by Jensen in his controversial (1969) monograph (p. 65), Herrnstein and Murray's figure still suggests a genetic factor that outweighs the combined influences of environment.

Among the small number of specific studies cited by Herrnstein and Murray to support their estimate, those involving twins are awarded special prominence. The authors distinguish between "direct" and "indirect" methods of assessing heritability, and declare:

> The purest of the direct comparisons is based on identical (monozygotic, MZ) twins reared apart, often not knowing of each other's existence.... The most modern study of identical twins reared in separate homes suggests a heritability for general intelligence between .75 and .80, a value near the top range found in the contemporary technical literature. (p. 107)

Because the most modern of the purest of the most direct estimates of heritability supposedly yields the highest values, Herrnstein and Murray's "middle-ground" figure of .6 is made to seem quite conservative.

In fact, however, their description can be challenged on several counts, starting with a first-hand examination of the main study to which they refer.

Raymond E. Fancher is a professor of psychology at York University, and Executive Officer of CHEIRON (The International Society for the History of Behavioral and Social Sciences). He is the author of *Pioneers of Psychology* and *The Intelligence Men: Makers of the IQ Controversy*.

The Minnesota Study of Twins Reared Apart (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990) reports raw intraclass IQ correlations of .69, .78, and .78 for 40-odd pairs of twins on what they designate, respectively, as Primary, Secondary, and Tertiary tests of intelligence. The study's authors, who scarcely can be accused of downplaying the influence of heredity in their overall interpretations, are nevertheless more conservative than Herrnstein and Murray in estimating heritability from these findings. They explicitly state that because very few of their twins had been reared "in real poverty or by illiterate parents," their estimate "should not be extrapolated to the extremes of environmental disadvantage still encountered in society." They conclude: "*In the current environments of the broad middle-class*, in industrialized societies, two-thirds of the observed variance of IQ can be traced to genetic variation" (Bouchard et al., 1990, p. 227, emphasis added). Not only is their "two-thirds" lower than the "between .75 and .80" proclaimed by Herrnstein and Murray, but it also is explicitly restricted to a "broad middle-class" population.

At places in *The Bell Curve* Herrnstein and Murray acknowledge that heritability coefficients pertain not to individuals, but to the variabilities of traits in specific populations, and explicitly note that heritabilities inevitably increase as populations become more environmentally homogeneous (pp. 106, 298). Clearly the main population with which they are concerned in their book extends from extreme to extreme of the range of socioeconomic status and is substantially more heterogeneous than the "broad middle-class" represented in the Minnesota study. Thus Herrnstein and Murray's claim that this key study points to a relevant heritability between .75 and .80, and thus defines the top end of their "competent and responsible" heritability estimates centering around .60, is mistaken and misleading.

Herrnstein and Murray's claim that separated-twin studies represent the "purest" of the direct measures of heritability is equally misleading. They do not even allude to the clear case made years ago by Kamin (1974) and Farber (1981), among others, that any separated-twin study can provide a "pure" estimate of heritability only to the extent that a number of important conditions have been met. The twin sample must be representative both of the general population whose heritability is to be estimated and of the subpopulation of all separated identical twins within that population. The twins must have been randomly placed in adoptive homes representing the full range of environmental advantage and deprivation. And the twins must have been reared truly separately, with no contact and ideally without any knowledge of each other's existence. Only to the extent to which these conditions have been met may the intraclass correlation between the adult twins' IQ scores (or any other measurable variable) provide a pure measure of heritability.

In fact, of course, a real study meeting these conditions might be done with fruit flies but not with human twins. Adoption agencies in real life inevitably and justifiably practice selective placement, and in the cases of twins make special efforts to keep them close to each other if not actually in the same household. Often they are placed in separate branches of the same family, with ample opportunities for contact. And of the small proportion of twins who have been truly and completely separated without knowledge of each other, only a few are likely to learn of each other's existence and identify themselves

for inclusion in a study. Those who do find out about each other usually do so via a common acquaintance who has remarked on their resemblance to each other.

In the late 1960s Cyril Burt's (1966) twin study briefly seemed the best in the literature because it included a table of the occupational levels of the adoptive parents for 53 separated pairs of identical twins indicating two things: The levels represented a full range from "Higher professional" to "Unskilled," and the intraclass correlation between the co-twins's adoptive parents' levels worked out to an astonishing –.03, indicating a virtually random placement (p. 143). Burt reported IQ correlations of .771 for "group tests," .863 for "individual tests," and .874 for "final assessments" (p. 146). Jensen (1969) gave this study particular emphasis when estimating IQ heritability at .8 (p. 52).

In the early 1970s, however, Kamin (1974) called attention to numerous deficiencies in Burt's report of his study. Crucial basic details were missing, including descriptions of the IQ tests employed, and even about the basic makeup of the sample (ages, sex of the twins, etc.). Unlike other major studies, Burt's included no detailed case reports on any of the twin pairs. Kamin further noted that as Burt had been periodically reporting on his study between 1943 and 1966 and the reported sample size increased from 15 to 21, to 42, and finally to 53 pairs in 1966, several of the reported IQ correlations remained constant to the second or even the third decimal place—a statistical coincidence far exceeding imaginable probability. Kamin concluded: "The numbers left behind by Professor Burt are simply not worthy of our current scientific attention" (p. 71). Jensen (1974) now conceded that Burt's data were "useless for hypothesis testing" (p. 24), and Burt's study has since been omitted from all serious discussions of the separated-twin data, including that by Bouchard et al. (1990).

Kamin (1974) went on to examine the three other major separated-twin studies published prior to 1974: Newman, Freeman, and Holzinger, (1937) had reported an IQ correlation of .67 for 19 separated twin pairs; Shields (1962) reported .77 for 37 pairs; and Juel-Nielson (1965) reported .62 for 12 pairs. In marked contrast to Burt's, these studies provided full information about all their twins, including not only the measures used to test them, but also details about their upbringing and their biological and adoptive families' backgrounds. Despite their interest, all these studies deviated significantly from the scientifically ideal model. The majority of twins had been placed in similar households (often separate branches of the same family) that had some degree (and often considerable) contact with each other as the twins grew up. Further, those twins who had been relatively most completely separated, and whose adoptive environments differed relatively most, showed the largest absolute differences in IQ: an unequivocal sign of environmental influence.

Farber (1981) carried this line of analysis farther, collating data from all 121 of the case studies of separated twins she could find in the professional literature (including the three studies cited above). Besides finding the pooled group unrepresentative of the general population (e.g., coming from biological families heavily skewed toward the lower end of the socioeconomic distribution), Farber (1981) reported:

> Only three [sets of twins] are pure cases who were separated in the first year, reared with no knowledge of twinship, and seen [by their investigators] at the period of first meeting. In other words, *of the 121 cases reported in the last fifty years, only three are "twins reared apart" in the classic sense*. (p. 60, emphasis in original)

While acknowledging the problems involved in pooling data from different studies, Farber (1981) conducted tentative analyses of the combined IQ data that suggested that between 20 and 25% of the variance was attributable to "factors associated with mutual contact" between the twins. This variance, she noted, is generally "hidden in the 'genetic' side of [most published heritability] estimates" (p. 206). Taking into account the degrees of imperfect separation in different ways and with different subpopulations, Farber recalculated heritability estimates that ranged from .14 to .67 (pp. 206-207). The point here is not to suggest a specific heritability figure, but rather to stress that the incomplete separation of "separated" twins in the literature prior to 1981 clearly had some effect in enhancing their IQ correlations, rendering those correlations overestimates of heritability to some degree.

Obviously Herrnstein and Murray's blithe description of "twins reared apart, often not knowing of each other's existence" (p. 107) fails accurately to characterize the samples in twin studies conducted before 1981. But what of the Minnesota study on which they place such emphasis? Unfortunately it is impossible to answer with certainty, because that study is reported only in a six-page article that contains none of the detailed case reports necessary to make an informed judgment. The report does include summary data, however, indicating that the twins had been together for an average of 5.1 months prior to their separation with a range from 0 to 48.7 months, and had had an average of more than two years' contact time with each other prior to their testing with a range from one week to more than 20 years (Bouchard et al., 1990, p. 224). Thus the Minnesota sample clearly shares at least some of the limitations of previous reputable studies, and it remains possible that once full details are published, independent reanalyses along the lines of Farber's and Kamin's will suggest IQ heritabilities even lower than the "two-thirds" proposed by the study's authors.

Although it is incidental to the scientific argument, a final word is in order about Herrnstein and Murray's brief and misleading treatment of the Cyril Burt "scandal." They first mention Burt (p. 11) as the subject of Oliver Gillie's London *Sunday Times* exposé, which alleged in 1976 that Burt's study of separated twins had been fraudulent. Without detailing the issues or any of the evidence involved, they proceed to a box entitled "The Burt Affair," which declares:

> It would be more than a decade before the Burt affair was subjected to detailed reexamination. In 1989 and 1991, two accounts of the Burt allegations, by psychologist Robert Joynson and sociologist Ronald Fletcher, written independently, concluded that the attacks on Burt had been motivated by a mixture of professional and ideological antagonism and that no credible case of data falsification or fictitious research or researchers had ever been presented. (p. 12)

Herrnstein and Murray do not mention Kamin's (1974) previous discrediting of Burt's study on purely scientific grounds, or Jensen's (1974) judgment that Burt's data were useless for hypothesis testing. (Kamin's book is cited, but

described only as a source of overheated political rhetoric against intelligence testing.) *The Bell Curve* neither discusses nor cites Hearnshaw's (1979) biography of Burt, which was written with full access to Burt's private papers and presented a detailed case indicating that Burt had no personal contact with twins nor with his alleged collaborators during the years his reported twin sample was growing so dramatically. Hearnshaw further documented numerous clearcut examples of unethical behavior by Burt; for example, he crammed the *British Journal of Statistical Psychology*, which he edited, with attacks on his critics actually written by himself but attributed to a large number of fictitious authors. Herrnstein and Murray do not mention that the "rehabilitation" efforts by Joynson and Fletcher are essentially critiques of details from the Hearnshaw biography, and that these critiques have themselves been subjected to serious criticism (Fancher, 1989, 1991; Lovie & Lovie, 1993; Samelson, 1992; Weizmann, 1992).

Herrnstein and Murray conclude their discussion of Burt on the following surprising note:

> An ironic afterword centers on Burt's claim that the correlation between the IQs of identical twins reared apart is +.77.... In 1990, the Minnesota twin study, accepted by most scholars as a model of its kind, produced its most detailed estimates of the correlation of IQ between identical twins reared apart. The procedure that most closely parallelled Burt's yielded a correlation of +.78. (p. 12)

There were two Minnesota measures that yielded correlations of .78—"a Raven [Matrices], Mill-Hill [Vocabulary test] composite," and "the first principal component of two multiple abilities batteries" (Bouchard et al., 1990, p. 224). But because Burt had described his measure extremely vaguely as "a group test of intelligence containing both non-verbal and verbal items" (Burt, 1966, p. 140), how can Herrnstein and Murray imply with such assurance that either of these Minnesota procedures "closely parallelled Burt's"? Herrnstein and Murray completely overlook the major point about Burt's study. What had originally set it apart was not the exact size of its correlations, for other studies had reported values in the same general range, including Shields's +.77. Burt's only really important claim was that his twins, unlike anyone else's, had been randomly placed in a representative range of home environments. *The Bell Curve* completely fails to indicate that no rehabilitation job on Burt so far, no matter how generously interpreted, has been able to produce credible evidence to support that claim.

In sum, the actual evidence from separated-twin studies offers a substantially less "direct" and "pure" estimate of IQ heritability than Herrnstein and Murray claim for it, and points to a value substantially lower than they assert. Because their "middle ground" heritability estimate of .6±.2 for the American population as a whole relies explicitly on separated twin studies to support the upper end of that figure, it is clearly an exaggeration to some degree. Had their consideration of separated-twin studies been truly fair and middle-ground, they might have concluded along the lines of Newman et al. (1937) at the end of their pioneering study almost 60 years ago:

> The farther one penetrates into the intricacies of the complex of genetic and environmental factors that together determine the development of individuals, the more one is compelled to admit that there is not one problem but a multiplicity of minor problems—that there is no general solution of the major problem nor even of any one of the minor problems.... We feel in sympathy [with the] dictum that what heredity can do environment can also do. (p. 362)

References

Bouchard, T.J., Lykken, D.T., McGue, M., Segal, N.L., & Tellegen, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250*, 223-228.

Burt, C. (1966). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology, 57*, 137-153.

Fancher, R.E. (1989). Fixing it for heredity. *London Review of Books, 11*(21), 19-20.

Fancher, R.E. (1991). The Burt case: Another foray. *Science, 253*, 1565-1566.

Farber, S. (1981). *Identical twins reared apart*. New York: Basic Books.

Hearnshaw, L. (1979). *Cyril Burt: Psychologist*. Ithaca NY: Cornell University Press.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement? In *Environment, heredity and intelligence, compiled from the Harvard Educational Review* (pp. 1-123). Cambridge MA: Harvard Educational Review.

Jensen, A. (1974). Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics, 4*, 1-28.

Juel-Nielson, N. (1965). Individual and environment: A psychiatric-psychological investigation of monozygotic twins reared apart. *Acta Psychiatrica et Neurologica Scandinavica*, Monographic Supplement.

Kamin L. (1974). *The science and politics of IQ*. Harmondsworth UK: Penguin.

Lovie, A.D., & Lovie, P. (1993). Charles Spearman, Cyril Burt and the origins of factor analysis. *Journal of the History of the Behavioral Sciences, 29*, 308-321.

Newman, H., Freeman, F., & Holzinger, K. (1937). *Twins: A study of heredity and environment*. Chicago: University of Chicago Press.

Samelson, F. (1992). Review of R. Fletcher's *Science, ideology and the media* and R. Joynson's *The Burt affair*. *Journal of the History of the Behavioral Sciences, 28*, 221-225.

Shields, S. (1962). *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press.

Weizmann, F. (1992). Cyril Burt, still lively after all these years. *Contemporary Psychology, 37*, 1331-1332.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 271-273*

David John Bateson
University of British Columbia

How Can *The Bell Curve* Be Taken Seriously?

How can a book that is so full of errors, so rife with inconsistencies and contradictions, and so blatantly biased and racist possibly be taken seriously at all?

One of my favorite little books of all time is entitled *How to Lie With Statistics* (Huff, 1954). This lovely piece of work coined the word *statisticulation*, which is defined as "statistical manipulation" or "misinforming people by the use of statistical material" (p. 100). The word is a good description of what is done in Herrnstein and Murray's *The Bell Curve*. Huff also points out that whenever data and statistics are presented to support any argument one should give the data and their treatment a critical second look. "Give that kind of second look to the things you read, and you can avoid learning a whole lot of things that are not so" (p. 19). This observation is appropriate for *The Bell Curve*.

When considering *The Bell Curve*, what amazes me is how a piece of work that is so poorly done, is so full of errors, is so rife with inconsistencies and contradictions, and is so blatantly biased racist could possibly be taken seriously at all. The fact that hundreds of media outlets (newspapers, magazines, radio programs, television shows) have discussed it, the American Educational Researchers Association/National Council on Measurement in Education annual meetings in San Francisco in 1995 contained a major, well-attended session on it, and a book presenting reactions by over 80 authors including Stephen Jay Gould, Irving Horowitz, David Suzuki, Arthur Jensen, and Carl Bereiter, to name but a few, has been produced (Jacoby & Glauberman, 1995) reinforces to me that the topic is an important one that elicits much emotional reaction. Obviously, people are concerned about the issues discussed in the book: the increasingly large and impenetrable barrier between the haves and the have-nots, the rich and the poor, the advantaged and the disadvantaged; the deterioration of what were once, and may still be considered, positive aspects of family and society; accelerating poverty, crime, and violence; education that seemingly does not challenge the bright or even constructively contribute to the "normal" student; and so on.

What concerns me is that such a poor, inferior piece of work that utterly lacks credibility might possibly have even a small influence on future social policy.

I am not a qualified sociologist, which tells me I should not expose my ignorance by attempting to discuss some of the social issues in the book that concern me. However, I do know something about statistics and data analyses,

David John Bateson is an associate professor in the Department of Curriculum Studies, Faculty of Education, and a research associate with the Educational Measurement Research Group.

logic, reasoning, argumentation, and methodology. It is in these areas that I make a few brief comments.

The problems that I found in the document are so numerous that it would take a book larger than the original to point out and explain them. Besides, many of these problems have already been exposed in numerous other situations as mentioned above. In fact, I doubt if there is a problem with the book that has not already been pointed out, so I will pay attention to only a few critical issues that can be used to illustrate what researchers should not do. The fatal flaws in the book, which the authors themselves point out, can become learning experiences for all researchers, novice or experienced; many of the mistakes are so easy to make when one is attempting to use research to "prove" one's point rather than using it to explore and learn about a question or problem of interest.

One of the good things about the book is that most of the substantial errors are already pointed out by the authors. I confine my comments to five major, fatal flaws that the authors themselves conveniently expose.

Probably the most striking flaw relates to the often improbable leap that these researchers (and I use the term loosely) make from uncovering a correlation to inferring, and then ratifying causality. All researchers are emphatically taught early in their development, and usually in their first statistics or methods course, that correlation does not necessarily infer causality. Many negative and often ridiculous examples are presented as illustrations in almost every statistics text printed. Taking a famous case as an example, one is not justified to conclude that churches cause crime just because there is a high, positive correlation between the number of churches per unit of population and the crime rate in urban centers in the United States. The authors of *The Bell Curve* themselves frequently and emphatically point out that one should not infer causality from correlation. They stress the point and repeat themselves often. After cautioning the reader, they then spend the rest of the book doing exactly what they have preached one should not do: they continually infer causality from correlations. This is one of the numerous contradictions and lack of coherence that permeates the text.

Another particularly glaring inconsistency is that early in the text the authors point out that using race as a category for discussion cannot be substantiated in any scientific way; that using race to distinguish among people cannot be justified in any way. They then proceed to use their category of "ethnicity" and to draw conclusions regarding and between their major subcategories of "black" and "white" throughout the rest of their book! A rose by any other name ...

A third awesome inconsistency is that the authors assert it is impossible to separate the relative contributions of genetics and environmental influences to what they call cognitive ability (rather than IQ). They acknowledge that, even within broad ranges, no satisfactory scientific proportions of these two influences have been even remotely confirmed. They then go on to conclude that environment contributes less than genetics and make the assumption that a "conservative" 60% due to genetics will form the basis of their arguments. From this they conclude that intelligence is "essentially" inherited through genes, and this then forms a fundamental cornerstone of their thesis.

They argue and base most of their conclusions and recommendations on the principle that cognitive ability, as measured by the tests they support, is essentially fixed, it is destiny, that nothing can be done to invoke change. At the same time, they report numerous counter-examples: "immigrants have sometimes shown large increases"; that poor education can lead to low scores; that children introduced into stable, supportive homes frequently show large gains; the list can go on and on. Strangely, they make clear the assumptions on which their arguments are based, and then show that these assumptions are patently false.

The final point I make has to do with data. Most researchers would heartily agree that the quality of any conclusions based on data is directly related to the quality of the data themselves. Based on their analyses of their data, the authors condemn the last 20 years of attempts to "equalize the playing field." Yet most of the key data that the authors use do not include results from any children who would have been influenced by any of the remedial, catch-up, enriching, desegregated, and so forth programs that have been introduced since the mid-1960s. The data the authors rely on most heavily tracked individuals through much earlier eras of development. However, they do not hesitate to use those data to argue against and even condemn recent programs, the possible effects of which the data do not and cannot capture. The lack of validity of their conclusions is astounding. How can anything so blatantly flawed be taken so seriously?

The five flaws above are so basic that they would not be accepted in even the worst graduate student paper; they are fatal weaknesses. Considering only these five problems, the entire thesis of *The Bell Curve* totally disintegrates.

Presented under the guise of being an objective, academically sound piece of work based on a long tradition of established research, the publication is so rife with spurious coefficients, weak links, biased use of data, misrepresentations, and blatant logical inconsistencies that it does not seem to deserve the attention it is getting. Perhaps it will serve the important purpose of bringing to the fore some of the immense problems our society presently faces. However, I doubt it. Instead it will probably do two things. First it will probably elicit destructive, emotional reaction from those opposed to blatant racism. The second, and probably most damaging, effect of the publication is that it may well discourage more ethical researchers from investigating some of the variables that require study for fear that they may be cast into the same pit of lions that Herrnstein and Murray deservedly have been cast by many who have critically examined their work.

Fortunately, most of the controversy and attention that the book has garnered have been confined to the United Sates of America. I would hope that Canada has enough sense to direct its energies to attempting to solve the pressing societal issues of today, rather than wasting energy in countering such meaningless drivel as presented in *The Bell Curve*.

References

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. Toronto: Free press.

Huff, D. (1954). *How to lie with statistics*. New York: Norton.

Jacoby, R., & Glauberman, N. (Eds.). (1995). *The bell curve debate: History, documents, opinions*. New York: Random House.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 274-276*

Parmeswara Krishnan
University of Alberta

The Bell Curve: Some Statistical Concerns

This voluminous book (Herrnstein & Murray, 1994), running into several hundred pages is packed with information, data, and the authors' theory that intelligence has a strong genetic component. My purpose here is not to comment on the rather ethnocentric views of the authors. My comments are confined to some of the methodological limitations of the research base of this book.

A Misnomer

The title of the book is a misnomer. The authors state often in the book that psychologists or psychometricians "make" the scores normally distributed. All the statements about the top, or the bottom, levels of students or people by IQ and other characteristics are based on the fit of the normal curve to the data. The normal (or the bell curve, or the Gaussian, or the error) distribution in its univariate or multivariate form can be generated in many ways. If IQ scores are a result of several additive forces (factors), then a normal distribution may ensue. The conditions for this have to be examined and analyzed. I would like to have seen some tests of goodness carried out before the blind use of the normal distribution.

Nonnormal Distributions

Several characteristics do not follow the normal distribution. Some are highly and some moderately skewed. Some have positive and some negative skewness. Some follow a Type I or a Type III Pearson distribution. Some follow the Pareto distribution.[1] Income in capitalist societies has a Pareto form. Studies done by Pareto himself and many others of income data for the United States show that a Pareto representation of the data is very satisfactory (Arnold, 1983; Lange, 1978). Studies done by Krishnan (1985) and Krishnan, Ng, and Shihadeh (1990) show that the generalized Pareto distribution applies to a wide range of social characteristics. Davis (1941), in his Theory of Econometrics, notes that the results of examinations in mathematics, the number of dissertations published in scientific periodicals, the efficiency in the game of golf and so forth. have a distribution similar to the Pareto distribution. These examples imply that further progress in the development of educational and physical efficiency are easier for persons who have already reached higher levels.[2] If abilities are Pareto in form and income also follows the Pareto curve, then the two are likely to be highly correlated and the joint distribution will be a bivariate Pareto.[3] The authors of the book have not looked at the various forms of Pareto distribution to develop their thesis.

Parmeswara Krishnan is a professor of sociology. His interests are in demography and mathematical modeling of social processes.

Use of Percentiles and Other N-tiles

The authors make copious use of percentiles and quintiles for making their points. These are fine if the data are normally distributed. For data on incomes and so forth, because the distributions are not normal the use of these n-tiles is suspect. Comparisons are possible with the help of some measures of inequality. The Lorenz curve or entropy or the simple index of dissimilarity would have been worthwhile. For instance, on pages 131 and 132 of the book, the percentages of people in poverty by parents' socioeconomic class and cognitive class are shown. The authors argue that as the cognitive level increases, the degree of poverty declines. If the index of dissimilarity is calculated, either for the poverty distribution or for its obverse, what we get is that the distributions of people by income and cognitive level are similar. The authors' conclusion is a bit hasty and rather simplistic.

Limited Multivariate Techniques

The authors use regression models. Other types of multivariate procedures could be profitably used here. For instance, Fisher's discriminant function can be employed (with a causal or noncausal framework) to distinguish between the rich and the poor, the whites and the blacks (or the nonwhites to include all other ethnic groups), the immigrants and the nonimmigrants. There are other techniques such as Mahalanobis D^2 statistic, general classification analysis, and Hotelling's T^2 statistic. Most of the regressions developed on pages 595-623 are poor linear fits. No meaningful inference can be made with these regressions unless nonlinear models are investigated. The modeling aspect is complex indeed. The authors should have been very cautious in their interpretation of the results.

Genetic Foreboding

The authors give a genetic slant to their results. They assume assortive mating of elite men and women. Is there a guarantee that the offspring of such matings will be bright or even brighter than the parents? In population (medical) genetics instances of inheritance of diseases in the mating of close kin relationships and of ethnic groups (e.g., Jewish group in North America) are discussed. The Mendelian laws yield us the ratios of the various phenotypes over generations. What the authors ignore are the possible social and family related consequences that may arise over time among such elite populations. Too much rigor and stricture on educational training may lead to deviant behavior in the young.

Final Word

This book has raised some important questions on the determinants of class structure in North America. It is left to the social scientists all over the world to refute or support the main thesis contained here based on rigorous statistical analyses.

Notes

1. The Pareto distribution or curve is given by the mathematical representation: $y = \frac{A}{(x-a)^{\alpha}}$, where x (horizontal axis) is the income and y (vertical axis) is the number of people whose income is more than x, 'a' is the lowest income at which the distribution (curve) begins and A and α are positive parameters. If we move the vertical axis to x=a, then

the Pareto curve takes a simpler form: $y = \frac{A}{x^{\alpha}}$ or $y = Ax^{-\alpha}$. The probability function of the Pareto distribution is $f(x,a,\alpha) = \alpha\frac{a^{\alpha}}{x^{\alpha}+1}$.

2. It can be shown that: $\frac{dy}{y} - \alpha \frac{dx}{x}$. This implies that the relative screening (decrease) in the number of persons as the income increases is smaller and smaller and diminishes in proportion to the income. The implication is that the advance to higher income class is easier for those who have already reached a high income than for those with low income.
3. Bivariates and multivariate Paretos are discussed in Arnold (1983) and Krishnan (1985).

References

Arnold, B.A. (1983). *Pareto distributions. Volume 5—Statistical distributions*. Burtonsville, MD: International Cooperative.

Davis, H.T. (1941). *Theory of econometrics*. Bloomington, IN: Indiana University Press.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Krishnan, P. (1985). Some demographic applications of the univariate and bivariate forms of the Pareto distribution. In *Proceedings of the International Statistical Institute, 45th Session, Amsterdam*. (pp. 267-268).

Krishnan, P., Ng, E., & Shihadeh, E. (1990). Some generalized forms of the Pareto curve to approximate income distributions. In *Proceedings of the Business and Economics Statistics Section, American Statistical Association, Washington, DC*. (pp. 502-504).

Lange, O. (1978). *Introduction to econometrics* (4th ed.). Oxford, UK: Pergamon Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 277-282*

John Gartrell
University of Alberta

and

Stephanie Amadeo Marquez
California State University, Hayward

The Spurious Relationship Between IQ and Social Behavior: Ethnic Abuse, Gender Ignorance, and Confounded Education

The Bell Curve is a pretentious, rather myopic study of stratification in the United States. Herrnstein and Murray (1994) argue that the expansion of educational opportunity in 20th-century America has produced a new "cognitive" elite. The good news is that a new more meritocratic educational recruitment process has realized the American dream of widespread educational and occupational opportunity. It has brought prosperity to many. The bad news is that the cognitively challenged have been left behind, and social welfare programs don't work because they fail to address this cognitive disadvantage. Indeed, it is hard to see what would increase cognitive abilities (beyond nutrition). Furthermore, because the well-to-do in the United States have stopped having kids, the burden of disadvantaged people will increase.

Has America come to be dominated by a cognitive elite selected in some sort of meritocratic process that stratifies society by intelligence? We think not. The ruling class of the United States (or Canada, for that matter) is not dominated by its advisers, or by the corporate lackeys who implement their will. The golden rule is still in force: "he who has the gold rules." Knowledge may be power, but you can buy knowledge, and knowledge is most effective for those who have the corporate means.

Is the structure of stratification in Canada and the US increasingly based in part on cognitive ability? To some extent it is. The reconciliation of these disparate answers lies in understanding the differences between the questions. The "brains trust" has not captured government and industry. The global village is not run by a cognitive elite, although it may be operated by one. The capitalist class has recruited a large number of clever helpers through an increasingly competitive educational system. The winners are well rewarded for their labor. If this is a "cognitive elite," then it works for a new capitalist class or for a more local national elite in a world capitalist economy. In a

John Gartrell continues his research on the use of positivism in sociology (with brother David), and on health in Native communities (with George Jarvis). He is just beginning a study of development and environment in Jasper National Park.
Stephanie Amadeo Marquez teaches sociology, women's studies, and criminal justice. Her current research examines gender and ethnic stratification, particularly as they are misrepresented in introductory social science college texts.

different sense, just because the important players in the capitalist world economy have received postsecondary education does not qualify them as a cognitive elite.

Of course, the contemporary trend is to try to make more money with fewer of these expensive "helpers." Downsizing and restructuring may be all the rage, but there is remarkably little indication that it guarantees profitability. It may undermine political stability if education is no longer viewed as *the* road to the good life. The meritocracy is being undermined. Many of the meritorious (those who receive bachelor's degrees, even professional degrees) are not rewarded with appropriate careers, reasonable security, and good incomes.

After all is said and done, how can Herrnstein and Murray expect to understand an elite whom they never study? Population statistics and random sample surveys are entirely appropriate to the study of the general population. They are not appropriate to study elites, except perhaps by comparison. Herrnstein and Murray's principle source of data, the National Longitudinal Study of Youth (NLSY), was designed to represent the general population. It is totally inadequate to study the capitalist power elite. They comprise but a few thousand families, and the probability of their being selected in a general random sample is remote indeed. Besides, we doubt that they would consent to participate, even if by some chance they could be contacted. If Herrnstein and Murray think that their cognitive elite is becoming isolated, they should try conducting interviews with really powerful people, even those in public office. Not only do Herrnstein and Murray use a methodology designed to ignore the elite, they also studiously ignore the voluminous sociological literature on social class, whether it is based on Weberian (Goldthorpe, 1987), Marxian (Wright, 1985), or completely different ideas concerning the labor theory of value (Sorensen, 1991). They also ignore the large body of economic and sociological research on education, as well as sociological research on social mobility or status attainment (for reviews see Kurtz & Muller, 1987; Wegener, 1992). We return to this literature because it so closely parallels their implicit models and makes many of the same mistakes.[1]

Leaving aside the ludicrous contention that intelligence rules America, or that power elites are selected primarily on their cognitive abilities, let us focus on the merits of the much narrower argument that educational and occupational attainment in America has become so much more meritocratic. What does this model look like? First of all, it is important to note that Herrnstein and Murray never make this model explicit. If they had done so, its central features would have looked just like the status attainment model outlined almost 30 years ago by Blau and Duncan (1967). The first part of the status attainment model represents mobility by identifying the effects of exogenous "ascribed" status characteristics of individuals (family of origin status, gender, ethnicity, and eventually IQ) as determinants of educational attainment. The second part of the model examined the effects of these exogenous[2] "ascribed" status characteristics and educational attainment on occupational prestige (both for entry jobs and subsequent occupational attainment).

This is essentially the same model that underlies Herrnstein and Murray's research. According to their argument, the cognitive elite is produced by a meritocracy in which cognitive ability is the principal determinant of educa-

tional attainment (particularly net of other ascribed status), occupational attainment, and many other things as well (occupational performance and income among them). Herrnstein and Murray do examine the effects of cognitive ability on attainment net of status of family of orientation. They demonstrate fairly convincingly, as have others before them, that cognitive ability is a stronger determinant of education than family of orientation socioeconomic status. Social status of family of orientation retains some influence net of cognitive ability, but we would except their argument that educational systems have become more meritocratic. We would accept the argument that gender and ethnic differences in gross educational attainment (years successfully completed) have decreased, and are relatively small once the influence of cognitive ability is taken into account. Differences in type of education remain.

The second link in the meritocratic chain is the effect of education on occupational outcomes (net of ascribed status). There are several serious problems with Herrnstein and Murray's treatment of this part of the model. First, in order to demonstrate that cognitive ability has a direct effect on occupational outcomes, they would have to control for the effects of education. This control is necessary to specify the indirect effects of cognitive ability through its influence on education, and any direct effects that cognitive ability might have on occupational outcomes, net of education. When they examine the effects of cognitive ability on occupational outcomes (including income), they do control for the social and economic status of family of origin. They do not control for education. In fact, they go so far as to use education as a replacement (their word) for cognitive ability in analyzing income differences (Appendix 6). As a result, they hopelessly confound the effects of education and cognitive ability. Their method of statistical analysis extends this problem to their observation of ethnic differences. They do not have this problem with observed gender differences only because they avoid gender other than to make it a control category.

The decision not to control for education while examining the effects of cognitive ability is rationalized (p. 124) by arguing, first, that they cannot control for educational attainment because it is dependent on parental socioeconomic status and IQ. Therefore, education "expresses" the effects of parental socioeconomic status and IQ. We accept their premise (parental status and IQ influence educational attainment). We do not accept their conclusion. Because IQ causally precedes education,[3] they must control for educational attainment when examining the effects of IQ on social behavior. Otherwise, they cannot distinguish the effects of cognitive ability (IQ) from those of education.

Their second rationalization for not controlling for education when estimating the effects of cognitive ability is that the effects of education are discontinuous. They point out that certification (e.g., high school diploma, college degree) may have an effect beyond years of schooling. This is a red herring. If they are correct, then the effects of education on occupational outcomes may be underestimated, because the variance associated with certification is not examined. This is hardly an excuse not to control for education at all when examining the effects of "cognitive ability." Rather, it is an effective argument for just such controls. Again, they demonstrate that they know how to overcome this kind of a problem. Indeed, the education variable that they use as a

"replacement" for cognitive ability in analyzing income differences (Appendix 6) is measured not in years successfully completed, but in five certificate levels. Besides, they repeatedly analyze specific certificate subsamples (high school diploma, university degree) when it suits their purposes.[4]

Their third excuse for not controlling for education when examining the effects of IQ is the possibility of multicollinearity. They imply that education and cognitive ability are just too highly correlated to both be used in analyses of occupational and other outcomes. Not one scrap of evidence is presented to substantiate this claim. They are very careful not to present the correlation between education years and the AFQT test (their measure of cognitive ability). They achieve this sleight of hand by introducing the NLSY only after they have finished arguing the connection between cognitive ability and educational attainment.

We doubt that the correlation between education and cognitive ability is so high as to introduce multicollinearity. As a rule, it would have to be well over .70. We suspect that they attempted a number of analyses in which both the AFQT and education years were included, and results did not support the argument that cognitive ability had substantial direct effects on occupational outcomes net of education. We suspect that its effects are in many cases primarily indirect through education, and because they are indirect, they are much smaller than would be necessary to sustain arguments about cognitive elites.[5]

Beyond these difficulties, Herrnstein and Murray abuse both gender and ethnicity. They employ these powerful, socially meaningful ascribed statuses primarily as control categories, rather than as interacting, explanatory variables. They tend to hide the main effects of ethnicity (Collins, 1989) and they manage to ignore the effects of gender altogether.[6] To make matters worse, the ethnic categories employed by Herrnstein and Murray are diverse. This makes it difficult to identify the meaning of ethnic differences without explicitly examining cultural characteristics of ethnic groups. East Asians may be Chinese, Korean, or Japanese, all with different cultures and similar enmities toward one another. Latinos include New York Puerto Ricans, recent Mexican immigrants to southern California, Cuban immigrants to Miami from the first wave of immigration, and sixth-generation Mexicans living in New Mexico and Arizona. With all this diversity, there is, of course, considerable variation in English literacy. Unless the AFQT was administered in "first" or "primary" languages, standard American English fluency would be hopelessly confounded with AFQT scores measuring cognitive abilities. Cognitive ability (thinking?) is thereby confounded with ability to communicate in English.

As if it were not bad enough that Herrnstein and Murray abuse ethnicity, they simply ignore gender. This demonstrates inexcusable ignorance. Although they do use gender as a control category (e.g., analysis of mothers of white children), most of the time they leave it out altogether (e.g., analysis of income). They justify their decision by strange logic, noting that there are no significant gender differences in educational attainment (years of education successfully completed) even within ethnic categories. This ignores differences in types of education (e.g., professional education), but that is not all. Lack of gender differences in education or AFQT scores (IQ) is hardly grounds to

assume that there are no gender differences in other "social behaviors," such as income.

In the US in 1992, women with college degrees earned $11,721 per year less than their male colleagues (Rothenberg, 1994). If there are no truly significant gender differences in cognitive ability, this large gender difference is not attributable to IQ. The consequences do not stop there. Differences that Herrnstein and Murray attribute to ethnicity or even to parental status may be attributable to gender; after reading their results we simply do not know. What we do know is that Herrnstein and Murray's observations as to the effects of IQ on social behavior are hopelessly confounded by both gender and education.

Herrnstein and Murray pull another statistical "fast one" by not evaluating how well their models fit in some cases (Appendix 6), or by creatively ignoring poor results when it suits them (dismissing poor model fits in Appendix 4). This statistical sleight of hand is used to exaggerate their claims for the confounded effects of cognitive ability on social behavior (chapters 5 though 12). The models in Appendix 3 that estimate the effects of AFQT on poverty (controlling for SES and age, not controlling for ethnicity or gender) both have a (pseudo) R^2 of less than .10. Less than 10% of the variance in poverty is explained by all the factors in the model, including IQ. In other words, 90% of the variance in poverty is left unexplained. This does not seem to bother Herrnstein and Murray. The detailed results for model estimates underlying Chapters 5 though 12 presented in Appendix 4 show that fully five out of eight of the more than 60 models account for less than 10% of the variance. In most cases, cognitive ability and all the other predictors explain remarkably few differences in social behavior (poverty, unemployment, idleness, injury, family matters, or parenting).[7]

Appendix 1: "Statistics for People Who Are Sure They Can't Learn Statistics." What marvelous irony. Statistical chicanery aside, we strongly suspect that the effects of education on occupational attainment (particularly income) have decreased during the 1980s and early 1990s. A university education no longer guarantees quality employment and the good things of life that secure incomes bring. Where does that leave the meritocracy? Where does that leave this new cognitive elite? Downsizing and restructuring in Canada and the US have clearly limited employment opportunities for college graduates (including professionals). The lucky, the connected, and most of the best may still get quality jobs, but this seems only to have exacerbated labor force inequality, particularly in Canada (Morisette, Miles, & Picat, 1994). What happens to political stability as the "cognitive elite" experiences unemployment and underemployment?

The Bell Curve does not address these issues. It can serve only as a very elaborate example for students of social stratification. Beware the spurious relationship. Beware the statistical lie. Accept no assumption unthinkingly. In the information age reanalyze the data yourself.

Notes

1. They do include Sewell and Hauser (1985) in their bibliography, as well as several other works. However, they do not make use of them in the text. In a spirit of evenhandedness, they ignore sociological literatures on social class and education, and economic research on income (and probably much else). This may explain why they miss one of the main criticisms of studies of attainment: that studies of occupational outcomes ignore those for whom wage

and salaried income is not the determining factor in their social status: the criminal, the homeless, and the very powerful or rich among them.

2. Exogenous factors are viewed as predetermined in the sense that they explain other variables in the model, but the model makes no attempt to explain them. They can, of course, be correlated. They may also interact as they influence educational and occupational outcomes.
3. They might also have argued that the direction of causation between IQ and education is difficult to specify. We would accept the data presented in Appendix 3 as clearly laying this issue to rest. Using independent data, they show that education has only small effects on later measures of cognitive ability.
4. They could have tried adding dummy variables representing thresholds in the education process to their regression models, or they could have assessed the utility of such a set of binary variables as a replacement for years of education.
5. For example, if (controlling for other factors in Figure 1) the net standardized effects of cognitive ability on education are .5, and the standardized effects of education on occupational prestige are about .5, then the indirect effects of cognitive ability would be .25. This would mean that for one standard deviation increase in cognitive ability, we would expect only one quarter of a standard deviation increase in occupational prestige.
6. For example, Herrnstein and Murray analyze "Latinos" and they also include gender effects in some of their analysis. They do not include Latino women as an analytical category. As a consequence, when they explain differences in occupational outcomes (e.g., income), they simply ignore gender differences.
7. The only models that seem to work reasonably well are those that use schooling and welfare as the relevant social behaviors.

References

Blau, P.M., & Duncan, O.D. (1967). *The American occupational structure*. New York: Wiley.

Collins, P.H. (1989). *Race, class and gender as categories of analysis*. Memphis, TN: Centre for Research on Women.

Goldthorpe, J.H. (1987). *Social nobility and class structure in Great Britain* (2nd ed.). Oxford: Clarendon.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Kurtz, K., & Muller, W. (1987). Class mobility in the industrial world. *Annual Review of Sociology, 13*, 417-442.

Morissette, R., Miles, J., & Picat, G. (1994). Earnings inequality in the distribution of working time in Canada. *Canadian Business Economics*, 2, 3-16.

Rothenberg, P. (Ed.). (1994). *Race, class and gender in the United States: An integrated study* (3rd ed.). New York: St. Martin's Press.

Sewell, W.H., & Hauser, R.M. (1985). *Education, occupation, and earnings: Achievement in the early career*. New York: Academic Press.

Sorensen, A.B. (1991). On the usefulness of class analysis in research on social mobility and socioeconomic inequality. *Acta Sociologica, 34*, 71-87.

Wegener, B. (1992). Concepts and measurement of prestige. *Annual Review of Sociology 18*, 253-280.

Wright, E.O. (1985). *Classes*. London: New Left Books.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 283-288*

Linda S. Siegel
Ontario Institute for Studies in Education

Does the IQ God Exist?

> The tendency has always been strong to believe that whatever received a name must be an entity or being, having an independent existence of its own. And if no real entity answering to the name could be found, men did not for that reason suppose that none existed, but imagined that it was something peculiarly abstruse and mysterious. (John Stuart Mill)

The fundamental premise of *The Bell Curve* is that a real entity, called intelligence, exists and is easily measurable. Herrnstein and Murray (1994) are true believers; they state, "That the word *intelligence* describes something real and that it varies from person to person is as universal and ancient as any understanding about the state of being human" (p. 1). In their world view, intelligence is accurately measured by a score on a test; this score is known as IQ (intelligence quotient).

What is Intelligence?

Herrnstein and Murray have no doubt that IQ really exists. They uncritically accept intelligence as an entity that is both real and measurable. This premise is incorrect and I present logical arguments and empirical evidence that contradict this assumption. I argue that IQ is merely a statistical fiction, an artificial construct that does not correspond to any real entity. In this commentary, I argue that as a concept, IQ is fundamentally flawed. However, IQ has a long and established history in psychology and challenging its validity is heretical. Let us begin the pilgrimage by considering some of the ways that Herrnstein and Murray conceptualize intelligence. They never actually define it (independently of a score on the IQ test) but provide some hints, based on historical sources, as to what they think intelligence might mean. For example, they note that Spearman defined intelligence as

> a general capacity for inferring and applying relationships drawn from experience. Being able to grasp, for example, the relationship between a pair of words like harvest and yield, or to recite a list of digits in reverse order, or to see what a geometrical pattern would look like upside down, are examples of tasks (and of test items) that draw on g as Spearman conceived of it. (p. 4)

The IQ score is a measure of *g*, and in a circular argument they assert that *g* is what the intelligence test measures. As Jensen (1980) has stated, "Intelligence is the 'g' factor of an indefinitely large and varied battery of mental tests" (p.

Linda Siegel, a Yale graduate, is a professor in the Department of Applied Psychology. She has conducted research in learning disabilities, language and cognitive development, the role of psychoeducational assessment in the identification of learning disabilities, bilingualism, premature and high-risk infants, and the early identification of learning disabilities. She is currently the editor of the *International Journal of Behavioral Development*.

249). Herrnstein and Murray believe that IQ tests measure *g*. As Jensen (1980) notes,

> We identify intelligence with "g," to the extent that a test orders individuals on "g," it can be said to be a test of intelligence. IQ is our most effective test of intelligence because it projects so strongly upon the first principal component (g) in factor analyses of mental tests. (p. 224).

Therefore, IQ tests measure *g*, and *g* is defined as what the IQ tests measure. Here we have a fundamental circularity from which there is no escape.

Typically, factor analysis is used to make it appear that the IQ score is an entity with an independent existence. A principal component factor analysis is used to analyze the IQ subtests and justify the existence of the IQ score as a real quantifiable entity, even though the principal components were arbitrarily obtained and vary for each individual. Thurstone's (1940) views on "g" are as valid today as they were then:

> Such a factor can always be found routinely for any set of positively correlated tests, and it means nothing more or less than the average of all the abilities called for by the battery as a whole. Consequently, it varies from one battery to another and has no fundamental psychological significance beyond the arbitrary collection of tests that anyone happens to put together.... We cannot be interested in a general factor which is only the average of any random collection of tests. (p. 208)

Gould (1981) cautions against the assumption that "test scores represent a single scalable thing in the head called general intelligence" (p. 155). He warns us to beware of the seductive statistical trap of factor analysis. The concept of *g* is based on a principal component (assumed to be *g*) emerging from the intercorrelation of subscales of an IQ test or IQ tests with each other. He notes,

> A factor analysis for a five x five correlation matrix of my age, the population of Mexico, the price of Swiss cheese, my pet turtle's weight, and the average distance between galaxies during the past ten years will yield a strong principal component. This component—since all correlations are so strongly positive—will probably resolve as high a percentage of information as the first axis in my study of pelycosaurs. It will also have no enlightening physical meaning whatever. (p. 155)

Multiple Intelligences

Another challenge to the concept of IQ is Gardner's (1983) concept of multiple intelligences, that is, the concept that intelligence is not a single entity but a complex set of mental functions that vary on several dimensions. This concept is dismissed by Herrnstein and Murray, primarily because the evidence for it is not of the statistical variety that is used to construct the mythical g. However, Gardner's basic point is ignored; human cognitive abilities are varied and rich and no paper-and-pencil tests can measure them accurately.

The Measurement of "Intelligence"

Another fundamental question is whether intelligence tests as they are currently constituted measure intelligence. The best way to approach this question is to examine the content of IQ tests. An examination of the content of IQ tests will show that they consist of measures of factual knowledge, definitions of

words, memory, fine-motor coordination, fluency of expressive language: however, these tests do not really measure reasoning or problem solving skills. They measure, for the most part, what an individual has learned, not what he or she is capable of doing in the future. Typical questions on the IQ test consist of definitions of certain words, questions about geography and history, tasks involving fine motor coordination such as doing puzzles, memory tasks in which children are asked to remember a series of numbers, and mental arithmetic problems in which the children must calculate answers in their head without the benefit of paper and pencil. Presumably, the basic aspects of intelligence are such abilities as logical reasoning, problem solving, and critical thinking. However, intelligence tests do not measure these abilities. It is obvious that these types of questions measure what a child has learned, not problem solving or critical thinking skills.

The bias inherent in the vocabulary questions of IQ tests is briefly discussed in Herrnstein and Murray (p. 281) with a good example of culturally biased material. Then this cultural bias is mysteriously dismissed with the wave of a hand (or a word processor). They argue that this item has the same relative difficulty for whites as it does for blacks (compared with other questions) and, therefore, is a valid item This logic escapes me. Instead, what it means is that this item is invalid for everyone because it depends on past experience or learning and not on some fundamental unlearned ability. Being able to answer these types of questions depends on the learning of some specific information. Therefore, the IQ test is a measure of what has been taught. Unless all children have the same learning experiences, differences in IQ scores measure differences in what has been taught, not some innate immutable quality.

Gould (1981) notes that Binet, generally acknowledged to be the father of intelligence testing, stated that intelligence cannot be measured directly in the same manner "as linear surfaces are measurable. If Binet's principles had been followed, and his tests used as intended, we would have been spared a major misuse of science in our century" (p. 151).

In addition, the IQ test is just a ranking of the individuals in a standardization sample who have taken the test. A serial ranking by definition cannot do more than rank order people relative to others; it is not a measure of absolute amount of individual capacity or potential. This simple concept of ranking has been misunderstood, misused, and confused with capacity. There exists no absolute measure of intelligence that can indicate potential, but only a serial ranking of performance on a group of tests called IQ tests. The misunderstanding of the nature and confusion of the IQ score as a measure of capacity may be due to the reification of the IQ score as a real and quantifiable measurable attribute.

IQ and the Individual Child

I would like to illustrate some of the problems with the use of IQ tests with a specific case, as outlined in Siegel (1990). This is the story of a real individual whose name has been changed to protect his privacy. Larry, age 8, received a score of 78 on an IQ test. He was placed in a class for mentally retarded children. This is a case in which the IQ score was used to make a decision that he was mentally retarded, and this fact should be kept in mind as the rest of the case is reviewed. He remained in classes for the mentally retarded until age 14.

Today, at age 34, he is enrolled in a graduate program in a major Canadian university after completing a BA in psychology with an A average. Larry had great difficulty learning to read, spell, write, and do arithmetic calculations. When tested at age 34 his IQ score was in the high average range; however, he still had significant problems with reading and spelling. His score on the reading (word recognition) subtest of the Wide Range Achievement Test (WRAT) was at the 18th percentile and his score on the spelling subtest of the WRAT was at the fourth percentile. His score on the Woodcock Word Attack subtest, a measure of phonological skills, was at the sixth percentile. He had difficulties on short-term memory tasks and had occasional difficulty with verb tenses and word finding in spontaneous speech. He had good general knowledge and vocabulary and an average score on a reading comprehension test. At age 34 Larry displays a profile of a reading disabled or dyslexic individual; yet at age 8 he was called mentally retarded.

Larry's case is a dramatic example of the consequences of using an IQ test score to makes decisions about an individual. At age 8, Larry was *reading disabled*, but instead was called mentally retarded. Larry was fortunate enough to have a determined personality and supportive parents who fought for his rights to be educated.

This case is a real one. Fortunately, it has a happy ending, but for many children with genuine learning problems the ending is not university or graduate school but jail, alcohol abuse, and/or suicide (e.g., Barwick & Siegel, 1995; McBride & Siegel, 1994). Larry's supportive environment did not prevent or cure his reading disability; his reading problem remained throughout his schooling and into adulthood. However, this environment probably prevented Larry from developing the serious social problems that are often a consequence of an undetected and untreated learning disability. Is Larry a rare exception? No. Today a child with poor reading skills and an IQ of 78 would be labeled "mentally retarded" or a "slow learner" or said to have a "general learning disability" and, in any case, would not be labeled as "reading disabled." Such a child would not be called learning disabled and would not receive intensive help with reading, because it would be argued, incorrectly, that we should not expect better reading from an individual with this IQ level. Unfortunately, children with low IQ scores who show signs of severe reading problems are still called mentally retarded even today.

Herrnstein and Murray might argue that this is only one case: IQ tests are not perfect, but in most cases they work. In fact, they have not presented any evidence of the validity of the IQ score for an individual case. They deal only with relationships among variables, for example, race and IQ, and provide averages, but do not show the validity of the measurement for a single case. Further, it is interesting to speculate that if IQ is not valid for an individual, it is really not valid for groups composed of a number of separate individuals.

I and others have argued that the IQ test is not valid for an individual case and should not be used to define a learning disability (for reviews and empirical evidence, see Siegel, 1988, 1989, 1992, 1995; Stanovich & Siegel, 1994, Toth & Siegel, 1994). It is often argued that we need IQ tests to measure the potential of a child. This type of argument implies that there is some entity that is real that will tell us how far a child can go, how much he or she can learn, and what we

can expect of that child. It is a logical paradox to use IQ scores with learning disabled children because most of these children are deficient in one or more of the component skills that are part of these IQ tests. Therefore, their scores on IQ tests will be an underestimate of their competence. It seems illogical to recognize that a child has deficient memory and/or language and/or fine motor skills and then say that the child is less intelligent because he or she has these problems.

The Stability of IQ Scores

Herrnstein and Murray argue for the stability of IQ scores except when they want to maintain (incorrectly) that IQ scores are not subject to improvement as the result of infant stimulation programs (p. 407). "To make things still more uncertain, test scores for children younger than 3 years are poor predictors of later intelligence test scores, and test results for infants at the age of 3 or 6 months are extremely unreliable." This statement is incorrect. They also note, "Up to about 4 or 5 years of age, measures of IQ are not of much use in predicting later IQ" (p. 130). This statement is also false; scores on IQ tests in infancy do predict later scores, but this relationship does not mean that IQ is genetically determined or impervious to environmental influences.

The IQ Religion

Herrnstein and Murray have built an elaborate structure and set of arguments that rest on the existence and validity of the IQ concept. If IQ is not a valid concept then this structure is doomed to collapse. Because IQ is not a meaningful concept, we are left not with an argument about the role of intelligence in determining the social structure of US society, but with a pile of rubble.

The answer to the question in the title is emphatically, "No, the IQ god does not exist." The elaborate structure that Herrnstein and Murray have built on the assumption that IQ is a measurable entity is as vulnerable as a sand castle is to the incoming tide. If IQ does not exist, then the rest of their arguments about class and racial and ethnic differences in IQ are meaningless. We need to beware of false gods.

Acknowledgment

The research described in this commentary was supported by the Natural Sciences and Engineering Research Council of Canada.

References

Barwick, M.A., & Siegel, L.S. (1995). *The incidence and nature of learning disabilities in homeless, runaway youths*. Unpublished manuscript.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gould, S.J. (1981). *The mismeasure of man*. New York: Penguin.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Jensen, A. (1980). *Bias in mental testing*. New York: Free Press.

McBride, H.D., & Siegel, L.S. (1994). *Learning disabilities and suicide: A causal connection?* Unpublished manuscript.

Siegel, L.S. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42*, 201-215.

Siegel, L.S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469-478, 486.

Siegel, L.S. (1990). IQ and learning disabilities: R.I.P. In H.L. Swanson & B. Keogh (Eds.), *Learning disabilities: Theoretical and research issues* (pp. 111-128). Hillsdale, NJ: Erlbaum.

Siegel, L.S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities, 25*, 618-629.

Siegel, L.S. (1995). *Adults with reading disabilities: The importance of phonological processing and the irrelevance of IQ*. Unpublished manuscript.

Stanovich, K.E., & Siegel, L.S. (1994). The phenotypic performance profile of reading-disabled children: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*, 24-53.

Thurstone L. (1940). Current issues in factor analysis. *Psychological Bulletin, 37*, 189-236.

Toth, G., & Siegel, L.S. (1994) A critical evaluation of the IQ-achievement discrepancy based definition of dyslexia. In K.P. van den Bos, L.S. Siegel, D.J. Bakker, & D.L. Share (Eds.), *Current directions in dyslexia research* (pp. 45-70). Lisse, Netherlands: Swets & Zeitlinger.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 289-296*

Robin Barrow
Simon Fraser University

Keep Them Bells a-Tolling

The fundamental argument about IQ testing, and claims about intelligence that are based to a large degree on research making use of such tests, is a philosophical rather than an empirical one. It is an argument about what "intelligence" means, whether *brain* and *mind* are synonyms or denote distinct concepts, whether it makes sense to conceive of an intangible, unobservable, and nonmeasurable capacity, or, conversely, whether it makes sense to imagine that there is some way of directly assessing intelligence. Some of those who employ standardized tests of intelligence as the basis of their research pay lip service to this point, though often it is only lip service. Others, despite over 100 years of intense discussion on the matter, seem quite oblivious of the fact that the major concerns are philosophical in nature and cannot, therefore, conceivably be laid to rest by any amount of further testing or statistical sophistication.

The popularity of Herrnstein and Murray's *The Bell Curve* (1994) is a little disturbing, therefore, given that, although they are fully aware that the value of their kind of research is disputed, and although they show some signs that they recognize that the issue goes beyond the technical proficiency of their work, they do not in the end give any indication that they understand the true philosophical nature of the problem. Indeed, the fairly lengthy bibliography rather gives the game away. There is no reference, for example, to Simon's pioneering (1971) *Intelligence, Psychology and Education: A Marxist Critique* first published in 1953 as *Intelligence Testing and the Comprehensive School*, no reference to the work of Ryle (1949), Glover (1976), Searle, (1984), or any other prominent philosopher of mind, no reference to Kleinig (1982), or any other philosopher of education (including myself, see Barrow, 1993) who has worked on the concept of intelligence, no reference even to those such as Young (1986), who while addressing the philosophical issues might nonetheless be sympathetic to the authors' position, nor finally to those psychologists who do show an awareness of the conceptual and logical problems inherent in intelligence testing such as Fontana (1981).

Herrnstein and Murray do at some level recognize that there is an argument that they need to face up to, but they have some difficulty in engaging with it. They know, for example, perfectly well that research of the kind they engage in ideally has to take account of all possible variables. They also rightly criticize fellow researchers who claim to be focusing on some new variable that is in fact merely a familiar one under a new name. But they seem neither to conceive of a number of particular variables that might affect test performance and have

Robin Barrow was a reader in philosophy of education at the University of Leicester before moving to Simon Fraser University where he is a professor and Dean of Education. He is the author of 21 books and numerous articles in philosophy, education, and classics.

not been researched, nor to recognize that the very idea that a research program could take account of all the variables involved is itself suspect to some. Basically, as they say, they are classical behaviorists, and as such they make the logical mistake of thinking that their theoretical perspective can be defended or proved sound by empirical research. They only have eyes for, they only believe in, what they can observe and measure. If one shares that premise, then most if not all of their argument is quite plausible, just as Herrnstein's syllogism to the effect that "if differences in mental abilities are inherited, and if success requires those abilities, and if earnings and prestige depend on success, then social standing will be based to some extent on inherited differences" is quite valid. But that's an awful lot of "ifs": the conclusion may be valid, but it may also be false. In the same way, for those who cannot understand why anybody should be so blind as to trust only their eyes, the behaviorist position is absurd, and would remain so regardless of what was contained in the 845 pages (including appendixes) of a book such as this.

Thus Herrnstein and Murray stress the degree of agreement among "the top experts on testing and cognitive ability" and, perhaps a little defensively, contrast academe with popular media pundits. But the problem is that the crucial questions are not the technical sort in which those referred to have expertise. So when, for example, they list six conclusions that they maintain are "beyond significant technical dispute," it is important to recognize that this may be true, without it being the case that the claims in question are true, have any real significance, or even mean anything coherent. Again, they know, of course, that "most mental tests are designed to produce normal distribution," but they don't seem to be sensitive to any of the implications of this fact: in particular, they seem not to realize that any conclusion drawn from such research to the effect that some people are more intelligent than others and most are about average is a foregone conclusion. One may accept that a high IQ is a good predictor of subsequent job productivity, higher educational attainment, prestigious occupation, and higher income, and that low IQ is a stronger precursor of poverty than low socioeconomic background, without being remotely impressed by the idea of using such tests. To accept that the IQ scores are good predictors is to say just that: generally speaking there is a high correlation. But, as the authors themselves are well aware, to say that it is generally the case tells us nothing about a particular case; there may well be numerous other factors that could decisively diminish the truth even of the generalization, and to make the judgment (whether revealed to the individual or others) may itself affect the outcome. But perhaps the main danger of such activity is that no matter what the testers say, IQ remains associated with intelligence in people's minds. And here we come to the nub of the matter.

The fundamental problem in the field of intelligence testing (and a great many other branches of empirical psychological inquiry into mental phenomena) is the lack of any coherent theory of intelligence (or mind) to give sense to and drive the practice. What I mean by a theory of intelligence in this context is a convincing articulation of the concept (which incidentally but importantly is not to be confused with a definition, even a clear one). I cannot here go into the question of what appropriate philosophical analysis of a concept involves (Barrow, 1991, Barrow & Milburn, 1990), but I think that understanding what

philosophical analysis is, and what it is not, could almost be treated as the hallmark of the bona fide philosopher; and, further, I believe that lack of this understanding is possibly the major problem that vitiates empirical research in the social sciences. Without any such articulation of the concept of intelligence, intelligence is inevitably defined by the instruments used to test it. We truly have not advanced beyond Boring's (1957) remark that intelligence tests only test intelligence, if intelligence is defined as "what the intelligence tests test." Any such definition, besides being prima facie unconvincing, would raise the question of why anybody should value or be interested in intelligence in this sense.

Much criticism of this book to date has centered either on technical discussion of the use and interpretation of statistics or on the fact that the authors make the "politically incorrect" or "socially awkward" claim that the IQ scores of American blacks are in general lower than those of American whites, and that this difference can be attributed to genetic factors. I shall ignore both these lines of criticism. It will be readily agreed that argument about the statistical procedures becomes irrelevant if the whole approach is misconceived, as I hope to show it is. The frightening idea that if something happens to be awkward or embarrassing we should suppress or deny it, even if we have reason to believe it is true, though hardly unexpected in these politically correct times, is too contemptible to waste time on.

There are, however, three devastating objections to the whole approach to intelligence involved in this research.

1. The normal curve or bell curve of distribution is not a discovery of the research or a conclusion to be drawn from it, but a premise (as the authors are of course aware). It is presupposed by all involved in IQ testing that intelligence will follow the pattern of distribution of something like height. It is visibly the case that a few people are short, a few tall, and most people distributed in the middle. It is hypothesized, for no very good reason that I can think of, that the same will be true of intelligence. Whatever its merits, this assumption remains a totally unproven hypothesis. But because the hypothesis is accepted, IQ tests when being designed are themselves tested for validity against this criterion. That is to say, a test that doesn't show a normal distribution is for that reason alone going to be rejected as invalid. Consequently, any researcher using standardized and accepted IQ tests is bound to end up with a normal distribution or bell curve.

 One reviewer, commenting on Murray's claim that the book's comments on blacks have been overemphasized, agrees, and remarks: "what the book really tries hard to prove is that most Americans—white and black—are not very bright, and that nothing much can be done to improve the bell curve distribution of intellectual ability" (Shelden, *The Weekly Telegraph*, No. 186). This is not perhaps phrased very fairly or felicitously, but what is of course true is that the main uncontentious conclusions to be drawn from the research are that IQ is distributed normally and that it is largely constant. What is also true is that the conclusions were essentially dictated by the mode of research (i.e., the nature of the tests). Something would be wrong if tests designed to show a normal distribution did not, and, a bit more subtly (see further below), something would be wrong if tests painstakingly

devised to exclude consideration of social or environmental conditioning were interpreted or seemed to tell us something about the external circumstances of the individual. Not that we aren't still left with rather a lot of questionable further claims, including the presumption that what seems to be innate, in the sense of not the product of external factors, is necessarily hereditary and the authors' view that 60% of IQ is inherited. But the main point to be made here is simply that the bell will go on tolling as long as people persist in pulling the rope; that is to say, we will learn nothing interesting about the *distribution* of IQ, so long as we continue to use IQ tests. We will simply affirm the preconception that its distribution follows the pattern of the bell curve.

2. The claim that whatever differences are recorded between individuals or groups are attributable to genetic factors is an inference that is likewise unproven. The claim gains what credibility it has from the fact that the tests are designed—albeit imperfectly, as we shall see—to test basic, rudimentary, and relatively limited or circumscribed skills that are not likely to be affected by social factors and, above all, education.[1] The reason for this is straightforward but once again self-fulfilling: it is presumed that there is something called intelligence that is in some sense located within the individual, which is perhaps to be distinguished from its developed state, which may owe something to external factors. Well, of course, if you treat intelligence as if it were an innate and fixed characteristic, like the color of one's eyes, and only test for something that is fixed or constant in the sense of unaffected by environment, upbringing or education, you will find something more or less constant. The question then becomes whether what you find or focus on bears any resemblance to what most people mean by intelligence or has any value.
3. The research imperative leads people to attempt to devise tests of skills that will not be affected by education or environment, even though that begs the question of whether these skills have anything to do with intelligence. However, as noted above, it is not in fact possible to do this job perfectly. Thus the differences between American blacks and whites, recorded by Herrnstein and Murray are not necessarily genetically based, because no test is likely to be completely culturally neutral and the question of whether one is or not cannot be determined from its use in research such as this. For example, one SAT item requires individuals to choose from four alternatives the pattern that is analogous to the pattern of "runner/marathon." The choice lies between (a) "envoy/embassy," (b) "martyr/massacre," (c) "oarsman/regatta," and (d) "referee/tournament." The testers proclaim the correct answer to be (c) ("oarsman/regatta"). It will be argued by many that this is a highly debatable contention. More to the immediate point, can one seriously maintain that there is no cultural bias here? Is everybody as familiar with the idea of regattas and oarsmen as everyone else? Was there never a person whose upbringing and background were such as to have left him or her unfamiliar with words like *envoy* and *embassy*? I cite this example because Herrnstein and Murray do themselves, precisely in order to deal with such criticism. But their discussion, while it fairly makes the point that as a matter of fact this item does not appear to discriminate against Blacks,

illustrates once again the great divide between a technical and a philosophical debate. No amount of evidence that as a matter of fact a given group does or does not appear to do as well on a specific test item as another group has any bearing on the question of whether the item is or is not culturally loaded. The question is whether it is coherent or plausible to maintain that understanding of analogous relations such as these is the product of some innate capacity rather than some aspect of upbringing. Put that way, it is difficult to see how anyone could offer an affirmative answer.

My argument, then, is that the entire research program carried out by Herrnstein and Murray is vitiated by the fact that it relies on tests that are designed to focus only on skills that are unlikely to be affected by time, place, or education and that are normally distributed. In fact the tests cannot be relied on to escape the influence of external factors, but at this point the question to emphasize is surely, assuming that the tests could be relied on to test only skills that were unaffected by external factors, why should we regard them as tests of intelligence, and why should we care?

There are only two ways in which such tests could be regarded as truly giving us information about intelligence: either we define intelligence in terms of the ability to do well on these tests, or we produce reason to suppose that good performance on the tests correlates with true intelligence. The first option strikes me as quite untenable: this is not what anyone means by intelligence, nor is there any particular reason to value the ability to display such rudimentary skills in and of themselves. The second option runs up against the uncomfortable fact that there is no such evidence. There is, it is true, some limited evidence suggesting a correlation between high IQ and subsequent academic achievement. But in a culture where academic achievement is itself to some large extent defined in terms of SAT scores, multiple-choice exams, and performance on tests of alleged generic skills of critical thinking, creativity and the like, this is to be expected. I am aware of no evidence to link IQ scores with academic achievement in the sense of ability to appreciate poetry, to write a respected history of some topic, to discourse intelligently on political theory, to solve the problem of Stonehenge, or to discover DNA. The Hawthorne effect may also be expected to play a part in the correlation between IQ and academic success. Individuals who are told (or treated as if) they have a high IQ might surely be expected to respond positively to this, whereas those who are told they have a low IQ are likely to become correspondingly demoralized. A similar line of reasoning must cast doubt on related claims about income, prestigious jobs, and so forth. The statistical data need not be in doubt, but there are a number of rather obvious reasons for expecting positive correlations between, for example, IQ, academic success, and well paid jobs. None of this gets us any nearer to judging a person's intelligence.

At this point we come full circle: one reason that there is no evidence of a correlation between high IQ as measured by standardized tests and intelligence is, of course, that there cannot be any such evidence in the absence of an adequate definition of intelligence in other terms. There have been various philosophical attempts to analyze the concept of intelligence (Barrow, 1993; Kleinig, 1982; Ryle, 1949), but these are routinely ignored by empirical workers in the field of IQ tests. If the philosophical work done on the concept of

intelligence were recognized, it would become immediately obvious that IQ tests do not measure intelligence and that there is no evidence and no obvious reason to suppose that there is any connection between IQ, as measured by standardized tests, and intelligence.

None of the above is to deny either the obvious truth that there is a neurophysiological basis to intelligence or the likely fact of genetic input. We know very well that, to take the extreme case, severely brain-damaged people cannot do certain things. The state of the brain, one may say, both limits and enables the development of intelligence. But the brain is to be distinguished from the mind, and it is the latter concept that is related to intelligence (Searle, 1984). The sad truth is that what Herrnstein and Murray see as a brave and heterodox championing of a generally condemned theory of intelligence is in fact mere clinging to an outworn methodological orthodoxy that betrays little understanding of the nature of the problem, which is essentially philosophical. Three of the six points they make at the outset of their book are:

1. There is such a thing as a general factor of cognitive ability on which human beings differ.
3. IQ scores match, to a first degree, whatever it is that people mean when they use the word "intelligent" or "smart" in ordinary language.
5. Properly administered IQ tests are not demonstrably biased against social, economic, ethnic, or racial groups.

The wording of each of these claims (and the specific reference to "beyond ... technical dispute" in the eyes of "top experts on testing") suggests that the authors take them to be empirical claims. Unfortunately, essentially they are not. The issue of whether there is such a thing as a general factor of cognitive ability is not the issue of whether, given that there are various general factors of ability, there is an intellectual one. The questions are what on earth is meant by a generic ability and whether it makes sense to think in terms of generic mental abilities (Barrow, 1990). Many powerful philosophical arguments have been adduced to suggest that it doesn't. But Herrnstein and Murray do not appear to be familiar with these arguments nor to appreciate that, whether convincing or not, arguments of this type are what need to be reckoned with. Similarly, it is hard to believe that IQ scores could match "whatever it is that people mean," given that people may mean different, even contradictory, things by "intelligent"; and the claim could not in any case possibly have been substantiated in the absence of a distinct articulation of whatever it is that people do mean. Finally, the claim that "properly administered IQ tests are not demonstrably biased," indicates yet again that the authors imagine they can settle the matter empirically. But the issue, as I have briefly illustrated above, is first and foremost one of considering the coherence of assuming that any given test item is supracultural. Whether one's understanding of oarsmen and regattas is in fact the product of one's social milieu is, of course, an empirical question. But the prior, nonempirical question is whether it is plausible to suggest that this and all the other elements like it in all the tests in use are somehow above and beyond consideration of class, race, gender, and so forth. It certainly isn't plausible to me.

None of the above implies that, on its own terms, Herrnstein and Murray's thesis is false, nor that some of the suggestions they make in the second half of the book are not sensible. Claims such as that "whites with IQs in the bottom 5

percent of the distribution of cognitive ability are 15 times more likely to be poor than those with IQs in the top 5 percent" and that "if tomorrow all job discrimination regulations based on group proportions were rescinded, the United States would have a job market that is ethically fairer, more conducive to rational harmony, and economically more productive, than the one we have now" may both be true. But the former is of no very obvious significance, and the latter neither follows from it (or similar claims relating to IQ) nor depends on it. Indeed, it is instructive that they remark of the latter claim: "We cannot prove that the proposition is true." But as a matter of fact they could, or at least they could attempt to, provided they understood that attempting to "prove" or "establish" such a claim requires philosophical argument rather than empirical data about IQ.

The difficulty the authors have in breaking free from the restraints of their scientism is illustrated well enough by some of their remarks on criminality. They are well aware that despite a strong correlation between low IQ and criminality, this does not mean that one should expect a person of low IQ to be criminal. As they honestly point out, "the great majority of people with low cognitive ability are law abiding." Furthermore, they concede, people are correct to maintain that by and large criminals come from "the wrong side of the tracks." But, they add, "the assumption that too glibly follows ... is that the ... social disadvantage is in itself the cause of criminal behavior." So far so good. They are right to stress that the correlation does not establish cause and effect. Indeed, in common with most researchers in this and similar fields, they repeatedly say throughout the book that a correlation is just that and is not proof of cause and effect. Alas, just like those other researchers, they also repeatedly draw inferences and make statements that *are* causal. Thus here they immediately continue: "That is not what the data say, however ... much of the attention now given to problems of poverty ... should be shifted to ... cognitive disadvantage." But there is no warrant whatsoever for the "should." The data themselves say nothing about whether we should focus on social disadvantage or cognitive disadvantage. They say only that many criminals come from the wrong side of the tracks and many of them do poorly on some rather peculiar, and in themselves trivial, tests.

Similarly, IQ testers need to give more thought to the point that there is not simply a problem of capturing all possible independent variables in one's research design, but a possibly insoluble problem in that the varying conjunction of discrete variables may itself be a variable. They also need to give more thought to the logical implications of the observation that in various respects test scores are improving. True, the commonsense position first advanced by Plato that one's intellectual performance was partly a product of heredity and partly environment allows one to say that any changes are due to environmental factors. But the more one recognizes that one can improve IQ scores, the less the allegedly innate factor would seem to matter.

I don't personally have any doubt that to some extent one's intelligence is a function of heredity, and I share the authors' view that to some extent we are simply socially afraid of asserting that some people are more intelligent than others. I also share their concern that "too few educators are comfortable with the idea of the educated person." However, the person who does well on IQ

tests has nothing necessarily to do with either the intelligent or the educated person, and, sadly, North America's longstanding obsession with trying to measure intelligence has itself played an enormous part in turning the school system away from any serious concern for educating the intelligent.

Note

1. There are, of course, some subscales that are consciously concerned with the effects of education. My point in the main text is related to the general desire (and need, given the perspective in question) to develop tests that are not affected by education.

References

Barrow, R., & Milburn, G. (1990). *A critical dictionary of educational concepts* (2nd ed.). London: Harvester Wheatsheaf Press.

Barrow, R. (1990). *Understanding skills: Thinking, feeling and caring*. London, ON: Althouse Press.

Barrow, R. (1991). *Utilitarianism: A contemporary statement*. Aldershot, UK: Edward Elgar.

Barrow, R. (1993). *Language, intelligence and thought*. Aldershot, UK: Edward Elgar.

Boring, E.G. (1957). *History of experimental psychology* (2nd ed.). New York: Appleton-Century-Crofts.

Fontana, D. (1981). *Psychology for teachers*. London: Macmillan.

Glover, J. (Ed.). (1976). *The philosophy of mind*. Oxford: Oxford University Press.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Kleinig, J. (1982). *Philosophical issues in education*. New York: St. Martin's Press.

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Searle, J. (1984). *Minds, brains and science*. Cambridge, MA: Harvard University Press.

Simon, B. (1971). *Intelligence, psychology and education: A Marxist critique*. London: Lawrence and Wishart.

Young, J.Z. (1986). *Philosophy and the brain*. Oxford: Oxford University Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 297-301*

Moya Kavanagh
University of Alberta

The Bell Curve: Rising Fears, Fallen Dreams

To begin, let me admit that much of the detailed evidence and argument presented in *The Bell Curve* (Herrnstein & Murray, 1994) is beyond my ability, or willingness, to appraise on scientific grounds. Though I maintain serious reservations concerning Herrnstein and Murray's conclusions about race and intelligence, I find their more general thesis describing the association between social behaviors and intelligence plausible. I wish to examine the implications of the latter with respect to a principle of equality; for I believe that the authors are largely correct in claiming that without instituting certain changes we will drift inexorably toward a custodial state. I also consider the extent to which particular changes the authors themselves recommend might correct such a drift.

Before I elaborate these ideas, I want to remark on my emotional response to the book. Without doubt, *The Bell Curve* is one of the most disturbing works I have read. While reading it, I often found myself thinking, "Oh, this must be wrong! This is terrible!" I want to try to elucidate what gives rise to this response, for I think I am correct in supposing that my reactions are not unique. As Herrnstein and Murray clearly realize, their theories evoke many concerns and fears. However, freedom is often fearful and painful. Intellectual freedom, if we are to exercise it responsibly, requires that we confront distressing issues and questions, and yet consider them fairly in the light of reason. What, then, are the specters that Herrnstein and Murray raise?

Perhaps the most haunting is the notion of the cognitive partitioning of society. A social hierarchy based on intelligence calls to mind such anti-utopias as Huxley's *Brave New World* (1977). More frightening still is the fact that Herrnstein and Murray's cognitively stratified society need not rely on sophisticated biological engineering, as Huxley's does. We need not tamper with developing embryos to ensure the production of alphas and epsilons in succeeding generations. The social and reproductive isolation of the different cognitive groups will successfully perpetuate the distinct classes. Furthermore, that an increasing proportion of freedom and power will reside with our alphas appears inevitable.

It is not only contemporary authors who have described societies stratified by intelligence. In Plato's *Republic*, Socrates describes the just state as one in which each man knows his role and keeps to it, be it the role of guardian, auxiliary, or artisan (Cornford, 1941). However, Socrates also tells us that each person is born with inherent attributes that suit him or her for one role rather than another. The myth of the metals, as it has come to be known, postulates different proportions of precious and base metals in the souls of different

Moya Kavanagh is a sessional lecturer in philosophy of education in the Department of Educational Policy Studies. Her areas of special interest are ethics and moral education.

classes of men. Gold, the metal that makes up the guardian's soul, is the metaphoric equivalent of intelligence. Plato's guardians are Herrnstein and Murray's cognitive elite.

It is one thing to entertain Huxley's and Plato's ideas, and quite another to consider Herrnstein and Murray's. Huxley's fiction is set some 500 years in the future. Plato's speculations were voiced 2,400 years in the past. Herrnstein and Murray speak of today.

Yet what is so worrisome about a cognitive elite? What disturbs so many of us about a society stratified by intelligence rather than title or wealth? It is simply this. Neither status nor wealth are heritable in the sense that intelligence is. The former are social constructs, dependent on the conventions of a society, and as such are subject to change. In contrast, intelligence is heritable biologically rather than socially. It comes with a passing on of genes, rather than the passing on of titles or estates. Hence a cognitively stratified society will be much more resistant to change, more firmly entrenched, more readily perpetuated.

Any social stratification is objectionable. Why? Because when we speak of a stratified society we speak of one with differential access to social goods, such as income, education, job opportunities, health care, and so forth. How much this access differs, and in terms of what particular goods, may vary. Hence I think that most of us would find Plato's republic a more desirable place to live than Huxley's brave new world. However, common to all such societies is the fact that the greatest power and freedom reside with the elite. As one descends the social scale, no matter how shallow the gradient, there is increasingly less choice and less chance to realize one's individual needs, desires, or interests.

If unequal access to social goods arises purely because some people merit the award of these goods though their own efforts, and if everyone has an equal opportunity to compete, then such unequal outcomes might be justified. In other words, if we start with a level playing field and let everyone play, perhaps we are right to consider the winnings fair. What Herrnstein and Murray make abundantly clear is the fact that the playing field is anything but level. Furthermore, given the largely heritable nature of general intelligence, it is difficult to imagine how we could level the field without resorting to Huxley's biological engineering to ensure that only alphas are born. Therefore, if we want a fair society, a just society, we must first recognize that it is not a level playing field. It is this fact that Herrnstein and Murray attempt to establish. Second, we must accept that the rewards of society, and access to social goods generally, are unfairly distributed; for such awards accrue to those who are merely born with high intelligence. In paying more for computer programmers than for window washers, for engineers than for plumbers, society rewards intelligence. However, we largely inherit intelligence; and for that part we don't inherit (our childhood environment) we are not responsible either. Thus to allocate social goods on the basis of intelligence is unjust. Although people may be accountable for what they do, given a certain level of intelligence, they are patently not responsible for their level of intelligence per se, and hence ought not to be rewarded or deprived of social goods on the basis of this criterion.

This is saying something more than "Life is unfair." What I am arguing here is that our allocation of social goods is unfair. Although there may very little we can do to rectify life's injustices, there is much we can do to amend society's.

This, then, is the specter that Herrnstein and Murray raise: a cognitively stratified society that, because of the heritable nature of intelligence, is remarkably fixed and resistant to change: a society, moreover, that is constructed to serve the interests and perpetuate the advantages of the cognitively elite. To the extent that disproportionate access to social goods accrues to this elite by virtue of a largely inherited trait, such a state is unjust. I find it repugnant and threatening, for I have no reason to believe that intelligence fosters virtue, or that cognitive ability increases the likelihood of a fair consideration of the interests of others. I am frightened by the thought that the foundations of this state are already firmly in place. So I can name my fears. I can understand why I found so much of *The Bell Curve* distressing to read. I am left with the question: what can be done? Have we any means to dismiss this haunting specter?

This brings me, though somewhat circuitously, to a consideration of the principle of equality. Many of the ideas I develop here can be found, in a more elegant and succinct form, in the work of Singer (1979). In the second chapter of his book *Practical Ethics,* Singer elaborates a principle of equal consideration of interests. It is in this respect, he argues, that all people are to be considered equal. Singer suggests that the important human interests include: "the interest in avoiding pain, in developing one's abilities, in satisfying basic needs for food and shelter, in enjoying friendly and loving relations with others; and in being free to pursue one's projects without unnecessary interference from others" (p. 21).

The first four interests appear to be independent of cognitive ability, inasmuch as they can be considered common to all, regardless of intelligence. The last of Singer's (1979) interests, however, will be influenced by intelligence. For example, the projects that an Einstein, a Wittgenstein, a Mozart might wish to pursue would differ substantially from those I would choose. So Singer's notion of human interests allows for individual differences. His principle of the equal consideration of interests simply prescribes an equitable provision for common interests and the freedom to pursue such individual projects as one's abilities and experience might engender.

With respect to individual differences and the diverse projects such differences imply, clearly Singer's (1979) principle entails equal freedom for all. Although my pursuits may not be interfered with, at the same time they may not impede the pursuits of others. Thus if one of my projects is the accumulation of sufficient wealth for a financially secure and comfortable life style, my means of realizing this end must not be at the expense of the interests of others. It is on these grounds, for example, that Singer illustrates the indefensibility of slavery. On these same grounds, we can clearly see that a state that rewards heritable intelligence with a disproportionate share of the social goods contravenes the principle of an equal consideration of interests. It is through our allocation of social goods such as income, education, and health care that we allow for the realization of important and common human interests. To the extent that we structure our society so that a certain group has greater access to

these social goods, to that extent we put the interests of the favored group before the interests of others.

On what basis could we possibly justify such inequity? Society rewards those who are lucky by birth, through the random combinations and permutations of the genes. At the same time, our social structure impedes the realization of the interests of those less lucky. This simply doesn't make sense to me.

Let us allow that a principle of the equal consideration of interests is an ideal toward which a society concerned with justice must strive. Given this, I wish to examine one or two suggestions that Herrnstein and Murray put forth in their final chapter.

First, it is important to note that there are intimations that the authors recognize some of Singer's human interests in their description of "a society in which every citizen has access to the central satisfactions of life" (Herrnstein & Murray, 1994, p. 551). They write:

> people can, through an interweaving of choice and responsibility, create a valued place for themselves in their worlds. They can live in communities—urban or rural—where being a good parent, a good neighbor, and a good friend will give their lives purpose and meaning. They can weave the most crucial safety nets together, so that their mistakes and misfortunes are mitigated and withstood with a little help from their friends. (p. 551)

It is clear from these and similar statements that the authors are concerned with the satisfaction of important human interests irrespective of intelligence or race. They recognize too that by focusing on individual rather than group attributes, we can remedy "the anger, the hurt, and the animosities" that otherwise continue to grow (p. 550). They cogently argue that it is within the context of local community, rather than government policy, that the individual is most likely to be treated as such. It is to the community, essentially the neighborhood community, that we must turn to effect the satisfaction of many human interests. Herrnstein and Murray claim:

> In a decent postindustrial society, neighborhoods shall not have lost their importance as a source of human satisfactions and as a generator of valued places that all sorts of people can fill. Government policy can do much to foster the vitality of neighborhoods by trying to do less for them. (p. 540)

I want to argue here that Herrnstein and Murray's focus on the individual and the local community can do little, given the present and predicted structure of society, to promote genuine equality of the consideration of interests.

First, consider their injunction to appraise people in terms of individual particulars, rather than in terms of the general group attributes, be the group racial or cognitive. Although this is necessary, it is not sufficient. In any society as stratified as ours, to ignore the group is to ignore an influence that strongly shapes individual merits and weaknesses. Because our society allocates social rewards differentially to groups, group membership affects the likelihood of an individual's acquiring particular traits and skills. For example, although intelligence may be heritable, learning to use one's intelligence, coming to know its limitations and developing its strengths, are not. Such learning, I suggest, is part of education: one of the social goods differentially apportioned to groups. Hence to promote strict individualism and to claim that "group differences can

take their appropriately *insignificant* place in American life" is to miss the forest for the trees (p. 550, my emphasis). By way of analogy, it appears that intelligence is determined by neither nature or nurture exclusively. In the same way, an individual person is the exclusive consequence of neither self-determination nor social determination. Both factors must be taken into account if we are to ensure an equal consideration of interests.

Second, Herrnstein and Murray's admonition that we look to community to realize many human interests is also unlikely to rectify present and future inequities. Again, given a stratified society and the physical isolation of different classes, it is apparent that certain neighborhoods will have more resources than others when it comes to satisfying human interests. We already recognize spatially localized populations as relatively homogeneous in many important respects, not least of which may be cognitive ability. Although people of any intellectual level may sometimes turn to friends and family for purpose, meaning, and support in their lives, the resources of some friends and families may be woefully inadequate for an effective response to these needs. Given today's level of social stratification, it seems to me unreasonable to suggest that communities can equally well promote access to "the central satisfactions of life" (p. 551). If stratification and class isolation continue, as the authors assure us will happen, how much less likely are we to see adequate communal support in less privileged neighborhoods? Hence reliance on community does little or nothing to alleviate the present or future unequal consideration of interests.

In the final chapter of *The Bell Curve*, the authors argue for a variety of policy prescriptions, of which I have examined only two. I suggest that it would be worthwhile to subject all their recommendations to careful review in the light of a principle of equal consideration of interests. Although space prohibits such scrutiny here, it is worth noting that Herrnstein and Murray allow that "Our views on all these issues are decidedly traditional" (p. 534). I think it fair to suggest that, as a consequence of their traditionalism, many of their suggested policies rest on unnecessarily narrow conceptions of equality, freedom, individualism, and community. However, I leave the justification of this claim to another time.

In closing, I believe the authors present their theories, their evidence, and their proposals in the genuine hope that such work may help ameliorate serious social ills. It may well do so, if only by virtue of the argument and debate it is bound to generate. Further, their descriptive claims, even if true, need not necessarily preclude a just society. What is required, however, is a careful consideration of their policy recommendations from a perspective that encompasses an equal consideration of interests. Without reformulation based on this principle, I fear that much of what they advocate will perpetuate and exacerbate the current injustices that characterize our society, to the detriment of all.

References

Cornford, F.M. (Trans.). (1941). *The republic of Plato*. London: Oxford University Press.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Huxley, A. (1977). *Brave new world*. London: Triad/Panther Books.

Singer, P. (1979). *Practical ethics*. Cambridge: Cambridge University Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 302-303*

W. Frank Epling
University of Alberta

The Bell Bottom

Herrnstein and Murray (1994) have produced a book that is concerned with intelligence, heritability, race, and social policy in the United States. The argument goes like this: blacks are dumber than whites who in turn are not as smart as yellows (but whites are only a little bit more stupid than Asians). You don't need to worry if you belong to a stupid race; if you are reading this journal you are probably OK. All that it means is that your race is dumb and you are a lucky exception. This can't be helped; it's a matter of your gene pool and it's not your fault. It's sort of like each race has only so much intelligence to spread around. When intelligence is doled out, some individuals get a bunch of it, but others only get a little bit.

Anyway, why make a big deal about intelligence? Normal and low intelligent people can have some good qualities. This is fortunate; otherwise you would end up with a bunch of dumb people who were really awful. Also, you can hate a smart person but cherish an idiot because of that individual's accomplishments and characteristics. There is even a small chance that intelligence is due to your environment. Although Herrnstein and Murray make these points, there are still reasons to be concerned about intelligence.

It seems that the more dimwitted the person, the larger, on average, is that person's family. We don't know why this is the case; maybe they have few interests and lots of time for purulent thoughts. Of course, deeply stupid people don't have large families because most of them can't figure out what to do to get a bunch of kids. This latter point is not directly addressed in *The Bell Curve*, but we can be reasonably sure that Herrnstein and Murray would agree that it is a fortunate fact about the profoundly stupid. When the incredibly dumb did manage to figure out the mechanics of reproduction they were mostly sterilized. Putting the issue of profoundly stupid people aside, it turns out that in the US the ratio of moderately dumb to smart people has been increasing: that is, the United States of America is turning into a nation of morons. Canadians have recognized this trend in the American population for some time, but have been afraid to point it out.

A Brief History of Intelligence

In the olden days of psychology, the French psychologist Alfred Binet constructed a test that was designed to identify areas of strength and difficulty for French schoolchildren. Many psychologists, however, reasoned that the test measured how intelligent the kids were, rather than what they had missed in school. Jumping ahead in the story of intelligence, English and American psychologists got real interested in stupidity. This was especially true for

Frank Epling is a professor of psychology. He has published several satires about psychology and about scientific racism.

American psychologists who were protecting the United States from a potential wave of dumb immigrants, and English educators who were required to separate dustwomen, janitors, and nurses from executives, engineers, and physicians. These psychologists needed reliable tests of intelligence and they made a whole bunch of them.

A major advance in intelligence occurred when Spearman found out that people really had intelligence. He found this out by using factor analysis. The location of intelligence had been suggested by Galton and others as more or less in people's heads. So by the turn of the century intelligence had been discovered and pinpointed. Once intelligence had been discovered a number of educators, psychologists, and others pointed out that it was inherited. This was good news because smart people could be mated to improve their race and dumb people could be sterilized. Also, this knowledge saved money, as it was obviously a waste of time and resources to try to boost the intellectual skill of poor and stupid kids.

Social Policy

Herrnstein and Murray rely on these facts; intelligence is a real physical entity that is passed from generation to generation. Presumably, if your mom and dad were a bit light of a full load, you stand a good chance of ending up on the dull end of the continuum. In modern America, social contingencies of reproduction favor dumb people and they are going to continue to increase. So what is the solution to this problem?

The solution is a typically American one. Return to the family values of old-time America as set out by the founding fathers of the country. Turn your .357 magnums and bazookas into friendly hunting rifles, get married and stay married, help your neighbor, just say no, join the Republican party, read *Readers Digest*, eat healthy food three times a day, and shake hands a lot. Once this is done, stupid people will feel good about themselves and they will get along better with smart people. All of this will help to restore people's dignity, especially if welfare is eliminated.

Reference

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 304-307*

Eric D. MacPherson
University of Manitoba

The Time Machine: The Next Generation

This attempt to circumnavigate *The Bell Curve* (Herrnstein & Murray, 1994) falls into three parts. The first has to do with the nature of social narratives (née social science) and the implications of those qualities for understanding *The Bell Curve*. Although there is an extensive literature of the genre, this essay draws particularly on Anderson (1981) and Postman (1992). The second, which builds on that foundation, refers mainly to H.G. Wells' (1895) *The Time Machine*, a remarkably prescient treatment of the theme of *The Bell Curve*. In the third I explain why the book ought to be burned.

What we have called social science can be laid out along a tapered strip or cone. At the pointy end we find questions to which we have secure answers. We have close to definitive ways to teach basic facts in arithmetic; we know how to find out, if an election were held today, how people would be likely to vote; we know which children's string figures can serve as cultural markers that can trace prehistorical migration routes in the Pacific; and so on.

A little farther along the strip we become less certain. We do not know what is the best way to teach long division or integer arithmetic; after a half-century of intensive efforts we are not certain as to how we ought to begin to teach children to read; and we do not know what the long-term effects of drug therapy are for highly active children. Questions of this sort are too convoluted, are too interactive, and entail too much feedback to allow for either reductionism or linear models. Instead of "answers," we must create what Cronbach (1971) and Maguire (1982) call *nomological networks*—syntheses of research, professional lore, and logical analysis that are probably better guides to practice than is street lore. Well along the strip we find what we will for now call large-scale social science—the zone of administrative "theory," "theories of learning," social "class," the Coleman Report, and the subject matter of *The Bell Curve*.

Anderson (1981), who should be cited more often than he is, summed up what had been decades of earlier criticism when he wrote that large-scale psychology is a *bogus science*. He shows clearly that no matter how we partition it, large-scale psychology does not satisfy any reasonable criteria for being considered a science, and to the extent that it pretends to belong in that company is a sham. We should regard it, he concludes, as a collection of belief systems. The same can be said of all large-scale social science. It is a collection of narratives, not science.

Jackson (1970) made the same point in what has become a classic underground reprint, "Is there a best way of teaching Harold Bateman?" The question, he argues, has no meaning. Like many others, Anderson and Jackson are

Eric MacPherson is Dean Emeritus of the Faculty of Education.

quick enough to point out the flawed conceptualization of large-scale social science, but rather short on recommended alternatives.

That is what makes Postman's (1992) essay noteworthy. He begins by reviewing its defects, and comes to the same conclusion; what we have called social science is, in fact, a collection of narratives—possibly instructive stories. The *New York Times*, he says, should cease reporting it on its Science page. But it is the remainder of his essay that matters. He rushes to make it clear that he does not mean that large-scale social science need be *only* storytelling. It can be a branch of moral theology, and as such can be most influential. It has had, and can have, major social effects.

> A novelist—for example D.H. Lawrence—tells a story about the sexual life of a woman, Lady Chatterley, and from it we learn things about the secrets of some people, and wonder if Lady Chatterley's secrets are not more common than we had thought. Lawrence did not claim to be a scientist, but he looked carefully and deeply at the people he knew and concluded that there is more hypocrisy in heaven and earth than is dreamt of in some of our philosophies. Alfred Kinsey was also interested in the sexual lives of women, and so he and his assistants interviewed thousands of them in an effort to find out what they believed their sexual conduct was like. Each woman told her story, although it was a story carefully structured by Kinsey's questions. Some of them told everything they were permitted to tell, some only a little, and some probably lied. But when all their tales were put together, a collective story emerged about a certain time and place. It was a story more abstract than D.H. Lawrence's, largely told in the language of statistics and, of course, without much psychological insight. But it was a story nonetheless. One might call it a tribal tale of one thousand and one nights, told by a thousand and one women, and its theme was not much different from Lawrence's—namely that the sexual life of some women is a lot stranger than and more active than some other stories, particularly Freud's, had led us to believe ... I do not say that there is no difference between Lawrence and Kinsey. Lawrence unfolds his story in a language structure called a narrative. Kinsey's language structure is called exposition. These forms are certainly different, although not so much as you might suppose. (p. 13)

And, toward the end of his essay,

> Like moral theology, social research never discovers anything. It only rediscovers what people once were told and need to be told again.... The purpose of social research is to rediscover the truths of social life; to comment on and criticize the moral behavior of people; and finally, to put forward metaphors, images, and ideas that can help people live with some measure of understanding and dignity ... [these] stories are rather more important than those of other academic storytellers ... We are obliged, in the interest of humane survival, to tell tales about what sort of paradise may be gained and what sort lost. We will not be the first to tell such tales. But unless our stories ring true, we may be the last. (p. 18)

Now, to some of the stories, particularly science fiction. The genre has attracted no more than its share of insightful tellers of tales, but since Ionian days those it has attracted have found its flexible parameters well suited to social commentary and speculations. At its best, the genre allows a few arguable or even improbable scientific premises, but from that point on, the science must be impeccable and the yarn is intended to work out the interaction

between that hypothetical world and the human condition. One of the first outstanding modern writers who exploited that freedom was H.G. Wells. The parallel between *The Time Machine* and *The Bell Curve* is very close, closer than even Postman would have thought possible.

Wells used a time traveller to project 800,000 years into the future his speculation that we are on the way to partitioning the human race into two species; the gentle, artistic and child-like Eloi, living in an above-ground paradise, and the disgusting subterranean Morlocks, the world's technical slaves. Wells has his time traveller point to the signs of the future split in contemporary London and adds,

> And this same widening gulf—which is due to the length and expense of the higher educational process and the increased facilities for and temptations towards refined habits on the part of the rich—will make that exchange between class and class, that promotion by intermarriage which at present retards the splitting of our species along lines of social stratification, less and less frequent. So, in the end, above ground you must have the Haves, pursuing pleasure and comfort and beauty, and below ground the Have-nots, the Workers getting continually adapted to the conditions of their labour. (p. 63)

Herrnstein and Murray tell the same story. It fully satisfies the criteria and is excellent science fiction. Beginning with one or two arguable but not improbable premises, they argue that since roughly the time of Wells we have sought to provide ways for the intelligent to rise through the social strata, and that more often than not it has worked. Far more than was the case in Wells' time or at any other time in history, they argue, the intelligent are now gathering above ground and the unintelligent are being sent below.

The parallels between the two stories continue. Wells' time traveller discovers a passageway to the lower regions and, out of curiosity, tries to ask the Eloi what is to be found below,

> They seemed distressed ... Apparently it was considered bad form to remark these apertures; for when I pointed to this one and tried to frame a question about it in their tongue, they were still more visibly distressed and turned away. (p. 62)

Just so. As long as Wells' world lay 800,000 years in the future, academics could be charitable and could perhaps even consider its biting commentary on contemporary England to be a cautionary tale. But when Herrnstein and Murray tell the story in expository rather than narrative language, discover the channels to the nether regions, and decide to ask questions about what is down there, we become distressed. It is bad form to "remark these apertures."

Writing reviews of *The Bell Curve* is a thriving cottage industry. One of the best is Bernstein's (1995) *The Poor Person's Guide to The Bell Curve* (presumably on account of the current hard-cover price of the book). He is the only reviewer I know of who admits to not having read the book on the grounds that doing so would spoil all the fun, but he is certainly not alone. Academics who do actually read all of it are in for a surprise. Given a few arguable but conventional assumptions, and supposing that it *is* science, it is almost a paradigm of how large-scale social science ought to be done; it is well "researched" and well balanced. Again and again, Herrnstein and Murray dismiss simplistic con-

clusions and call attention to varied possible interpretations of their massive data. And yet Wells is right. Academics in the main are "visibly distressed and turn away." And they do so because Postman is right. It has little to do with science; *The Bell Curve* is bad moral theology.

We are obliged to tell tales about what sort of paradise may be gained and what sort lost; help people live with some measure of understanding and dignity. And on those grounds we must turn away from *The Bell Curve*. One problem is that the cognitive-ability filter has not been so effective as Herrnstein and Murray may believe. Few reviewers seem to have got to the last few pages in which they point their moral, that we ought to tailor our social interventions to a new reality, and there are ample signs that there are larger numbers who are prepared to take the tale as an excuse to further disengage polite society from the Morlocks; to gradually deprive them of dignity, decent schooling, health care, and adequate nutrition.

We therefore have good reason for preferring Gould's (1981) narrative *The Mismeasure of Man*. Better to be told that we may be able to breed dogs for intelligence and temperament, but not people, and better to believe that given the opportunity virtually any child can become whatever he or she pleases.

So what to do with *The Bell Curve*? Burn it. Or at least do to Herrnstein and Murray what we did to Arthur Jensen: dismiss them as misguided academics who are playing out an outdated and simplistic theory of intelligence. Their story, and H.G. Wells', may well turn out to be prescient, but we have a better chance of avoiding that fate if we consider too close an examination of the Morlocks to be in bad taste and turn away.

That's how moral theology works.

References

Anderson, C.C. (1981). The bogus science. *Canadian Journal of Education, 6*(4), 8-22.

Bernstein, J. (1995). The poor person's guide to *The Bell Curve*. *Scientific American, 272*(4), p. 140.

Cronbach, L. (1971) Test validation. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.

Herrnstein, R.J., & Murray, C. (1994). *The bell Curve: Intelligence and class structure in American life*. New York: Free Press.

Jackson, P. (1970). Is there a best way of teaching Harold Bateman? *Midway, 10*(4), 15-28.

Maguire, T. (1982, June 9). Stating the art of measurement: The shadow recommendations. Paper presented at CERA, Ottawa.

Postman, N. (1992). *Conscientious objections*. New York: Vintage Books.

Wells, H.G. (1895). *The time machine*. Reprinted 1968 as *The time machine/The war of the worlds*. Greenwich CT: Fawcett Books.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 308-311*

Rebecca Priegert Coulter
University of Western Ontario

And Academic Sexism Too: A Comment on *The Bell Curve*

It is not unusual for individuals to assert scientific objectivity as a cover-up for a specific ideological position and to use purportedly scientific data to suggest or defend a particular type of social policy agenda. Gould (1981), Sayers (1982), Tavris (1992) and many contributors to Harding (1993) demonstrate quite clearly how science has been used to support, among other things, the argument that social inequalities have a biological basis or are genetically inherited and thus natural, right, and good. And, as McWilliam (1987) points out, "it follows that those seeking a greater measure of 'equality' are seeking to alter what is natural" (p. 67). Herrnstein and Murray's (1994) book *The Bell Curve: Intelligence and Class Structure in American Life* is simply part of a continuing historical pattern in which science is called into service for racist and sexist purposes.

Although critical reviews of *The Bell Curve* have focused on its shaky claims about race, heredity, and intelligence, little attention has been given to the sexism implicit in much of the book. However, like the "scientific" approaches of craniometry and eugenics, Herrnstein and Murray's methods and arguments reveal not only fixed attitudes toward African-Americans, but a deep-seated ideological position on the role of women. Indeed, in what reads like a naive and simplistic faith in intelligence testing that could easily be dismissed if it were not so dangerous, the authors sound remarkably like their 19th- and early 20th-century predecessors. The craniometrists, however, actually measured something that could be measured, namely the size of the brain, although they were mistaken in their interpretation of what brain size meant. Through their science, they established that the brains of non-Europeans were smaller than those of Europeans. Because brain size was equated with mental ability (the larger the brain, the brighter the person), it was argued that non-European groups were mentally inferior and thus would actually benefit from subjugation by European populations. This, then, became the biological justification for slavery and colonization. When it was also established that women's brains were smaller than men's, scientists and physicians, among others argued that this established women's mental inferiority, which because it was "natural" was not amenable to change through education. Women's smaller brain size, then, was used to explain why educational reform would not be of any use in correcting women's lack of intellectual achievement and the policy consequence was to deny women access to better and higher forms

Rebecca Priegert Coulter is an associate professor in the Division of Educational Policy Studies, Faculty of Education. Her research interests include gender and education and the history of youth.

of education. In this tradition, Herrnstein and Murray's ranking of humans by cognitive groups leads to similar policy conclusions about education. Low cognitive groups, which coincidentally are dominated by African-Americans according to the authors, are not likely to benefit from educational interventions, and indeed funding should be channelled away from programs for the disadvantaged to programs for the gifted. Of course, Herrnstein and Murray try to soften the implications of this position by acknowledging that some individuals in all groups are in the high cognitive range, but the force of their argument is clearly that we should divert educational resources to the privileged and accept that "many people will not reach the level of education that most people view as basic" (p. 436).

To give Herrnstein and Murray credit, they do not suggest that women's intelligence is inferior to men's. However, their views on women's role in society is quite clear. Women belong in the home raising children, and men belong in the work force. They speak of the existence of the traditional nuclear family as "a universal law." Citing the early 20th-century work of Bronislaw Malinowski, they reassert his claim that a group consisting only of the mother and her offspring is "sociologically incomplete and illegitimate" (p. 177). This one source is used to justify the labeling of children born to single mothers as "illegitimate" and the ferocity and lack of feeling with which they attack "illegitimacy" establishes a clear line to the policy position taken by elements of the radical right in the United States. In fact, Herrnstein and Murray argue that the policy intervention most likely to work in raising the cognitive ability of the population as a whole is adoption at birth from a bad (i.e., single mother) family environment to a good one. They explicitly want to turn the clock back to the period before the 1960s when single, pregnant women for a time gave up their children for adoption at birth. There is an obvious contradiction here between their argument that intelligence is largely inherited and this one that clearly gives a large role to environment.

Herrnstein and Murray argue that marriage should be returned "to its formerly unique legal status" and that "marriage once again become the sole legal institution through which rights and responsibilities regarding children are exercised" (p. 545). Because it is not at all clear why, using their terms, two low-intelligence parents are any different from one low-intelligence mother, except as an ideological principle, and why the presence of low-intelligence fathers (men from the group given to more violent and criminal behavior according to the authors) should mediate neglect and abuse in the home, one can only conclude that Herrnstein and Murray's real agenda is reinforcing the patriarchal family. In their policy prescription, Herrnstein and Murray also suggest that unmarried mothers should have no rights in claiming child support from the putative father, a view that continues their line of reasoning that it is women who are responsible for pregnancies out of wedlock and women who are responsible for parenting. That is, in all their discussions of "illegitimacy," birth control, and parenting, it is women who are the center of attention. In fact, a reader could be left with the sense that immaculate conception is the main mode of reproduction among the low cognitive "underclass" in the US. Men appear miraculously to receive full attention in discussions of employment and crime.

Despite their occasional bouts of denial, Herrnstein and Murray take a punitive and controlling approach to people they perceive to be outside of the cognitive elite, and their take on science, their shaping of data, and their policy recommendations are strikingly like those of the eugenics movement of the earlier decades of this century. Indeed, the parallels in language and in policy prescriptions are rather remarkable. One of the leading eugenicists in Canada during the 1920s and 1930s was Madge Thurlow Macklin. A medical scientist and social conservative, she argued that the major social problem facing Canada was the rapid growth of a group of feeble-minded, mentally ill, and defective people in the population. These people were a drag on the taxpayers, tended to marry thoughtlessly, had large numbers of children they could not support properly or, even worse, they were sexually promiscuous, producing large numbers of defective, illegitimate children. Hence the defectives and feeble-minded were passing on tainted genes at a rapid rate and, because of their low mental abilities, were such poor parents they abused and neglected their children. In contrast, good middle-class families, those that were independent, hard-working, thrifty, virtuous, and intelligent, were limiting family size and as a result the overall intelligence of the population was declining. This was made even worse, Macklin argued, by an immigration policy that did not police the mental capacities of those seeking new homes in Canada (McLaren, 1990). With little variation, this is precisely the line of argument taken by Herrnstein and Murray, although they finger African-Americans for particular censure.

Although the eugenicists spoke quite openly about restricting the fertility of women of low mental abilities through sterilization, and Herrnstein and Murray's work will clearly lead many to the same conclusion, they do not go so far as to suggest this policy solution. They are forced to settle instead for what they clearly regard as weaker solutions. Painting a soft, romantic picture of a nearly forgotten time of small communities and neighborliness, they suggest efforts to force a return to the traditional family and what they see as the old middle-class values of hard work, diligence, virtue, and monogamy. They claim to want everyone to find a valued place, a place where "other people would miss you if you were gone" (p. 535), even if you were not very smart. They want to simplify life for the dull so they can cope more readily. But then they get down to their hard-nosed policy recommendations. They target poor women and children whom they argue are of low cognitive ability, by urging the elimination of the social welfare system and "the extensive network of cash and services for low-income women who have babies" (p. 548). They support programs that would encourage poor women to use birth control through easy access to safe, inexpensive, and foolproof methods. They also believe that "competency rules" should be established for immigrants and that family reunification should not drive immigration policy. Although Herrnstein and Murray make no mention of the fact, such a provision has significant gender and race components.

Like the craniologists and eugenicists before them, Herrnstein and Murray use science to argue that social divisions and cognitive partitioning cannot be stopped or changed because they are natural, based on genetics and the heritability of intelligence. In other words, a society structured by class and

race, by differential access to wealth and privilege is meant to be. This cannot be changed, but at least the lower cognitive order can be provided with bread and circuses in their own neighborhoods while males in the high cognitive group (Newt Gingrich?) get on with running the country and the world and perhaps figuring out how to ensure smart women have more babies in order "to raise the IQ" (p. 548) of American society.

References

Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.

Harding, S. (Ed.). (1993). *The "racial" economy of science: Toward a democratic future*. Bloomington, IN: Indiana University Press.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

McLaren, A. (1990). *Our own master race: Eugenics in Canada, 1885-1945*. Toronto: McClelland & Stewart.

McWilliam, E. (1987). The challenge of the new right: It's liberty versus equality and to hell with fraternity! *Discourse, 8*, 61-76.

Sayers, J. (1982). *Biological politics: Feminist and anti-feminist perspectives*. London: Tavistock.

Tavris, C. (1992). *The mismeasure of woman*. New York: Simon & Schuster.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 312-315*

Ruth Reviere
Dalhousie University

"Slow Through The Dark": A Commentary on *The Bell Curve*

I was in the Caribbean last fall (1994) when I first heard about a new book that was attempting to revive the old and hurtful stereotype of black intellectual inferiority. Friends of mine who had read about the book and its conclusions regarding race and IQ in the October 24 issue of *Newsweek* drew it to my attention. My friends' reactions ranged from mirth to incredulity. You see, in the Caribbean we are somewhat shielded from this insidious debate, so it is hard to accept that approaching the 21st century as we are, there are still scholars who are willing to believe and propagate these centuries-old and discredited notions of black intellectual inferiority. I, however, have been living in Canada since 1991 and so have had time to appreciate that Herrnstein and Murray's ideas were quite likely to reach fertile ground. The fact that by the end of 1994 the book had sold nearly half a million copies and had reached number 5 on the New York Times Bestseller List only proves my point.

I returned to Canada in November 1994 to find Charles Murray, the surviving member of the pair of authors, the darling of the media and consequently being provided with unlimited opportunities to propagate his views. That Murray's views have found such a receptive audience, although disappointing, is not at all surprising for, as Hacker (1992) argued, white people have an unhealthy fascination with race and racial differences especially as these pertain to black and white people. The fact that coverage of the book virtually ignored comparisons of cognitive ability between whites and Asians, and between classes, and focused almost exclusively on black-white differences confirms my point.

I must confess that as a black person and scholar, I have had about as much as I can take of these hateful ideas that keep resurfacing century after century, essentially unchanged, by cowardly racists hiding behind the so-called objectivity of science. We all know that scientists, even those of the so-called cognitive elite like Herrnstein and Murray, do not live in a social vacuum. They are as easily influenced by the values and assumptions of society as the rest of us. We all know that race is a hot topic, and race combined with intelligence is an even hotter one. Herrnstein and Murray knew this too, and that is why the book was written the way it was. If the authors had confined their analysis to class, although their findings would have been essentially unchanged (class and race in America being so inextricably linked), they would not have generated the interest and hence the sales that they did. Herrnstein and Murray do not come over as the dispassionate scientists that they claim to be but rather

Ruth Reviere is a doctoral candidate. Her research interests include physics education, IQ and race, and Afrocentric research methodologies.

as cynics who willingly exploit a regrettable flaw in the American culture for their own ends.

This book has been critiqued elsewhere extensively, and justifiably, on its theoretical and methodological distortions and inaccuracies. This special issue, I am sure, will contain much more of the same. However, as an Afrocentric scholar I find it unproductive to critique comparative research, especially that comparing racial and ethnic groups, solely in terms of scientific validity, especially when science is used merely as a buttress for offensive opinions and to provide an aura of legitimacy and respectability. Instead, I want to comment on three issues. The first is the lack of respect for truth that the authors display throughout this book. The second is their willingness to encourage disharmony between groups they identify as cognitively different. And the third is their lack of interest in questions of basic justice.

Lack of Respect for Truth

Herrnstein and Murray have subverted the quest for truth, which ought to be the goal of all good researchers, to their transparent ideological aims. They attempt to disarm the reader by stating anticipated criticisms of their inferences, thereby giving the illusion of even-handedness, and then continuing blithely along their own ideological path to their plainly erroneous conclusions. Let me illustrate with a few examples.

Example 1. Despite ample evidence to the contrary, Herrnstein and Murray dismiss environmental influences on IQ scores with the mind-boggling assertion that over the last 30 years all Americans, regardless of race or class, have been provided with equal opportunities because they all have the same educational experiences. They argue that environmental effects are therefore canceled out, leaving only genetic differences to determine success. This, surprisingly, comes after a declaration early in the book that they were agnostic on the issue of whether IQ variations were caused by the genes or the environment.

Example 2. Early in the book, they insist that IQ is independent of educational attainment, but in the introduction to Chapter 12 they refer to education as the "proxy" of IQ and go on to draw conclusions about the effect of IQ from correlations with education. Throughout they refer to correlations of numerous variables with IQ, and draw causal inferences from these (in the introduction to Part III they go so far as to describe their aim as that of demonstrating causality), even though they themselves point out that causality cannot be inferred from correlations. As they claim correctly in the introduction to Chapter 16, "causal relationships are complex and hard to establish definitely" (p. 369).

Example 3. Herrnstein and Murray come to the unsubstantiated conclusion that "the success of the early waves of West Indian blacks seems unlikely to repeat itself." They go on to explain that

> In his book *Ethnic America*, Thomas Sowell described the successes of West Indian black immigrants, starting from early in the 20th century, noting among other things that, by 1969, second-generation West Indian blacks had a higher mean income than whites. (p. 363)

Bearing in mind that they argue that earning potential correlates positively with "cognitive ability," rather than coming to the logical conclusion (based on their own argument), which incidentally I do not subscribe to, that this must therefore mean that, at least in 1969, West Indian black immigrants were more intelligent than white Americans, this example of black success is dismissed out of hand as an aberration. The reasoned explanation, of course, is that immigrants, regardless of nationality, race, or creed tend to be highly motivated and oriented toward success.

Example 4. Herrnstein and Murray cite an attempt in Venezuela, among other examples, to increase IQ by special coaching and report that an increase of 0.4 SD occurred for a "conventional intelligence test." They acknowledge that "there was no chance to see if the gain faded out or was reflected in the rest of the students' academic performance, nor can we even guess how much a second or third year of lessons would have accomplished" (p. 400). Despite this, on page 402 they conclude that "the goal of raising intelligence among school-age children more than modestly, and doing so consistently and affordably, remains out of reach."

Disharmony Between Groups

Through this book Herrnstein and Murray cynically attempt to encourage disharmony between groups they define as cognitively different. They clearly wish to foment fear and distrust of an alleged underclass that they portray as dangerous, unrestrainable, and totally beyond help. It is not coincidental that this underclass comprises poor people, with Africans and Hispanics heavily overrepresented. These are the people popularly perceived as disadvantaged, and the authors' goal is to convince their readers (the supposed cognitive elite) that disadvantage with its attendant social ills is preordained, and that this underclass poses a direct threat to the well-being of the cognitive elites. I cite a few examples of how they do this.

Example 1. Herrnstein and Murray open with the formidable promise that "this book is about differences in intellectual capacity among people and groups and what those differences mean for America's future" (p. xii). They go on to talk of intelligence as a "force for maintaining a civil society" (p. 254) and to identify those who possess this commodity (the cognitive elite) and those who do not (the underclass). In their vision of America, everyone will be confined to a "valued place" to be determined by their IQs. A warning runs throughout the book that in order to have a stable society the cognitive elite must be protected from the consequences of the low intelligence of the underclass. The authors come across as unabashedly racist, classist, and elitist, and I say this without apology, for they articulate pretty strong views about the inferiority and superiority of different groups of people and advocate that people be treated in different ways because of the group to which they belong.

Example 2. Herrnstein and Murray claim that there are too many blacks on college campuses in relation to their IQs, especially in the elite colleges. They suggest that blacks are taking the places of better qualified whites because of affirmative action and that this causes whites on campus to view blacks as intellectually inferior. They argue, improbably, that "American universities once [during the mid-1950s to mid-1960s] approached the ideal in their handling of race on campus and there is no reason why they could not do so again"

(p. 476). This is an astounding contention, for this decade describes the period immediately following the Brown decision on school desegregation when there was tremendous racial conflict on college campuses with the National Guard escorting black students onto campuses to protect them from the rage of white protestors.

Basic Justice

Herrnstein and Murray seem totally unconcerned about issues of basic justice. Their agenda is clearly ideological with the acknowledged intent of influencing social policy, and they are dismissive of concerns about social justice. Here are some examples.

Example 1. The authors rail against corrective measures based on group membership while they themselves define the problem in those terms. If, as they claim, an individual's performance is explained by his or her membership in a specific group, then it is unfair to argue against the solution being similarly race-based. It is worth noting, however, that they do not have a similar compunction when it comes to punitive measures, and advocate that these measures should be targeted directly at the offending populations as they have defined them. For Herrnstein and Murray, you reward individuals but you punish groups. A genuinely individualistic approach would be inimical to all race-based analysis, including IQ.

Example 2. Herrnstein and Murray warn repeatedly about the power of the "cognitive elite" to subvert democracy, but if they were really concerned about this danger they would encourage a greater representation and not fight against the presence of blacks and Latinos among the Ivy League sets. The racist attitudes they suggest are present among the cognitive elite—and which, they warn, may be more evident and bolder in the future—are hardly a manifestation of intelligent behavior. But they do not condemn this failing. They describe a sort of Nazi/police state that may be the future of America, which they suggest may be an inevitable response by the cognitive elite to the presence of blacks, poor whites, and Latinos among them. The fact remains, however, that if the cognitive elite were ever to arrive at a position where they could put this threat into effect, they would have to be judged by the same standards that others who perpetrate crimes against humanity are judged. The perceived intellectual disadvantages of the so-called underclass cannot be used to justify such an atrocity.

In conclusion, I maintain that Herrnstein and Murray are not the dispassionate and objective scientists they claim to be. No such being exists, especially when it comes to such a deeply personal issue as IQ and race. The issue that confronts me as a black scholar is how do I deal with such material as *The Bell Curve*? I must confess that I do not yet have the answer. I do know, however, that I am past anger. Now I feel only despair. When will it end?

References

Hacker, A. (1992). *Two nations: Black and white, separate, hostile, unequal.* New York: Ballantine.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 316-321*

Erin Gibbs Van Brunschot
University of Alberta

and

Augustine Brannigan
University of Calgary

IQ and Crime: Dull Behavior and/or Misspecified Theory?

The Bell Curve (Herrnstein & Murray, 1994) has been met with the perennial torrent of criticism that greets every attempt to reduce complex social behavior to a biological basis. The reduction to genetics of important personal attributes shared across specific racial and ethnic groups courts charges of racism. But even if the book's basic premise about genetic differences proves to be incorrect or overstates the importance of genotypes, the fact remains that profound differences exist in American society in areas related to academic performance, early school leaving, and delinquency. If political conservatism makes such differences appear to be natural and intractable, we should be equally suspicious of the ideological presuppositions of learning theorists who imply that the only impediments to social achievements are those external to individuals. Neither extreme is realistic.

Increasingly, there is a recognition of the need to develop interactive models that reflect the impact of factors from various levels (Gove, 1994). In the past, proponents of IQ causation (i.e., Goring, 1913) have estimated that the correlation between mental defectiveness and criminal behavior was an impressive –.65! Most criminals were idiots by this estimation. Theories of hereditary feeblemindedness and policies of sterilization and selective immigration were hastily adopted to prevent what Herrnstein and Murray (p. 343) call *dysgenesis*—the erosion of national levels of intellectual prowess. These monocausal models and the worrisome policies they abet have led criminologists to discount completely the role of IQ in delinquency: another type of extremism.

In the latest debates, researchers have argued that IQ is a better predictor of delinquency than either class or race, a situation that ought to be greeted with enthusiasm to the extent that it suggests that careers in delinquency ought to be understood on the basis of personal characteristics. However, *The Bell Curve* is explosive because it implies that IQ is an aggregate racial trait and that it is heritable. Because, the story goes, nurture differences have been leveled by

Erin Gibbs Van Brunschot is a doctoral candidate in sociology. Her research interests include methodology, deviance, and crime, and her recent research has examined prostitution and assaultive behavior.
Augustine Brannigan is professor of sociology at the University of Calgary. His recent research has examined problems of street children and prostitution in Canada.

Head Start programs, welfare and affirmative action, the relative standing of social groups that survive must arise from nature. This joint reification of intelligence and race is social dynamite, for it grounds the social hazards of poverty in biology. But neither race nor intelligence is immutable. Racial characteristics and boundaries are subject to constant movement, and aggregate levels of intelligence on a national level have moved as much as 15 points in this century (the Flynn Effect). Even if IQ is 90% heritable, that would leave sufficient room to explain IQ differences on the basis of phenotypical starvation. Our position recognizes that there are differences in the IQ scores of delinquents and nondelinquents whether estimated on official data or self-report studies. However, although the relationship between IQ and delinquency may well be real, it does not necessarily have the origins or effects suggested by Herrnstein and Murray. In this discussion we note that the effects of IQ on delinquency are mediated by gender and age in a fashion that is not readily explained by a reductionist perspective, and that the relationship between IQ, school performance, and delinquency may be spurious.

Gender and Crime

The authors limit their analysis of the relationship between IQ and crime to white men by admitting that "crime is still overwhelmingly a man's vice" (p. 245). In their analysis of the National Longitudinal Survey of Youth (NLSY), having found a negligible but statistically significant relationship between criminality and IQ among white men, one would expect to draw the same inference with respect to women. This conclusion is not borne out by studies of female delinquents (Townes, James, & Martin, 1981). Balthazar and Cook (1984) found not only that there was no relationship between female violent crime and intelligence, but also that there was no relationship between female intelligence and incarceration. Notwithstanding problems with the use of incarceration as a measure of crime, the analysis employed by Herrnstein and Murray is not supported for females who would appear to be more criminogenic than males, given their lower IQ scores. The negligible findings of the relationship between male violent offending and IQ, and the lack of a female IQ-violent crime relationship, suggests that either a different theory is required to explain female offending (which is in itself problematic) or that IQ does not hold the explanatory power attributed to it.

Females are not, however, left out altogether. Low intelligence appears to have conflicting effects depending on gender. For males it appears that low IQ predicts nonconformity, whereas low IQ for females predicts conformity (Naffine, 1987). Herrnstein and Murray's argument draws on stereotypical notions of the "nature" of women versus the "nature" of men. Women's crimes or wrongdoings have to do with reproduction and dependency, that is, becoming unwed mothers and going on welfare, whereas men's crimes arise from aggression and agency, hence the focus on violent crimes.

The idea that low IQ could cause such vastly different reactions in males and females subverts the theoretical importance of the IQ-crime linkage. Empirically we know that women are less likely to be in public situations conducive to violent crime either as victims or perpetrators (Sommers & Baskin, 1993). Simpson and Elis (1995) also note that "crime, as social action, is a 'resource' for accomplishing gender—for demonstrating masculinity within a

given context or situation. Thus prostitution draws on and affirms femininity while violence draws on and affirms masculinity" (p. 51). As well, differential sources of control, that is, formal control for males, and informal control for females, also imply that IQ has little to do with criminal involvement. These observations suggest that criminal behavior is at least as dependent on opportunity and socialization as on IQ.

Age and Crime

Herrnstein and Murray suggest that IQ tends to be relatively fixed over the life cycle. Earlier, Wilson and Herrnstein (1985) had argued that "the early onset of misconduct is one of the best predictors of a child's becoming a chronic and persistent delinquent," and "measures of aggression are almost as stable over time as measures of intelligence" (p. 242). It is surely ironic how much Herrnstein's rather brittle "learning theory" has come to reflect 19th-century ideas about the "born criminal": neither appears to accord much value to experience, nor much change arising from it. In point of fact, delinquent behavior is notoriously age sensitive. Among young men, whose intelligence is supposedly fixed, delinquent conduct rises sharply in the midteen years and declines in the early 20s. This appears to be the case across different cultural and temporal contexts. Furthermore, although most serious adult criminals were young offenders, the vast majority of young offenders do not become adult criminals. Hirschi and Gottfredson (1983) argue that it is age per se that predicts delinquent activity, that age is not a spurious measure or a proxy for a better specified cause, and that the age of first delinquency is indicative of nothing later (except to say that those people who commit a lot of crimes as adults did the same thing when they were youngsters). In other words, the tendency to engage in delinquency varies tremendously over the life cycle, whereas intelligence is supposedly fixed. Obviously the correlation between delinquency and IQ is secondary to gender and age factors. This is reflected in the evidence Herrnstein and Murray present from the NLSY.

Herrnstein and Murray employed two measures of crime: (a) self reported crime and (b) evidence of incarceration based on interviews conducted in a correctional facility. If we examine (a) in their "basic analysis" (n=2,008), we find that the model tested included the effects of IQ, SES, and age. Notably, age was a significant predictor (b=–.203) as was IQ (b=–.269). The original NLSY sample included respondents aged 14-22 in 1969—a range that captured only the rising change in crime and age, and not the downturn expected as youth reached their mid-20s. In the NLSY, age is as substantively relevant a correlate of delinquency as IQ. So far, so good.

In the High School Sample (n=665), IQ became nonsignificant and age remained a significant negative correlate of delinquency (b=–.261, α=.07), as we would expect following Hirschi and Gottfredson (1983). And in the College Sample, none of the coefficients was significant, probably because neither age nor delinquency varied enough to make the test meaningful. Even allowing for the reliability of these tests, the explained variance was pitiful: 1.5% for the first model, 2% for the second. In other words, although the coefficients were significant, their explanatory power was trivial.

As for dependent variable (b), interviewed in jail, the overall models improved in terms of explained variance (9%), but the only variable that proved

significant in the basic sample and the high school sample was IQ. The strong negative coefficient (b=–.63, α=.0001, n=1,945) for IQ suggests that these respondents were probably highly skewed in IQ. This may occur because very low IQ correlates with incarceration, a factor one would find in populations that are not just deficient in IQ but more liable to detection, arrest, and detention as a result. In addition, IQ performance may be partially related to social class, and the resources of social class may confer an ability to evade criminal processing. Here the relationship could be indirect and still significant. In that interpretation, IQ may be conflated with SES, and the formal 5-point scale used to test SES in *The Bell Curve* is simply too crude to disentangle the two.

Last point, detention in jail, although equated with serious criminality, has other causes. It might apply to persons who do not qualify for pretrial release because they have no fixed address (i.e., the homeless), because they are in breach of conditions of probation from previous convictions, or because they are serving time as a result of a new felony and may have found themselves behind bars, or because they were determined undeserving on extralegal grounds (i.e., unemployment or gender). Of what is IQ indicative? Homelessness, insobriety, unemployment or crime? *The Bell Curve* suggests criminality, an inference that may be sound but hardly compelling in view of equally plausible alternatives. But even if we allow the correlation, what can be said of the linkage?

Education, IQ, and Crime

There are different ways to make sense of the relationship between delinquency and IQ. Herrnstein and Murray pointedly assign IQ as a causal factor: "Clearly something about getting seriously involved in crime competes with staying in school" (p. 250). The burglar probably "is not the most obedient of pupils" and the street fighter "probably gets in fights on the school grounds." School and delinquency are incompatible, and low IQ begets school failure, begets crime. Wilson and Herrnstein (1985) suggest that "it is in school that a youngster lacking intellectual skills is most likely to encounter for the first time the frustrations and failures that his or her deficits are heir to" (p. 168). Intellectual failure, we are told, makes "short term horizon" delinquent behavior ("self-indulgence") more attractive. But this line of reasoning merely defines delinquency circularly with attributes of low IQ, as if reading well was not its own form of immediate gratification that did not indulge the self. The notions of self-indulgence and immediate gratification are used to discount immoral activities that are done quickly and effectively without regard to the Protestant ethic. The myth that doing what is "good" is necessarily difficult is just that, a myth.

An IQ-delinquency model not considered by Herrnstein and Murray is the "school reaction" model, whereby "IQ may influence delinquency indirectly because it may be used as a criterion for differential treatment in the schooling process" (Ward & Tittle, 1994, p. 190; Menard & Morse, 1984). Essentially, children have many resources that they can use to maximize positive feelings and to control their environments. Boys who are larger than their classmates can use physical leverage, not by actually being physically aggressive, but by acting in a dominant way. They can also develop verbal skills and friendship networks as sources of power. But not everyone has the same resources

(Agnew, 1990). Mesomorphs—the so-called body type associated with knuckleheaded delinquents and brawny football players—are probably more likely to get away with roughhousing than 90-pound ectomorphs. We label mesomorphic power pejoratively as "impulsive" and "aggressive," though "initiative" and "leadership" might be more positive ways of characterizing the same behavior. If institutions like schools remove children who are more physical, who "act out" more, who test the status hierarchy by fighting, then such children will attract more social control, expulsion, and negative labeling in schools and, at least initially, will do so independently of their scholastic abilities. In other words, their conduct may be labeled delinquent because it is institutionally inappropriate however much it has been successful in prior family and preschool situations.

There are other problems with the relationship between IQ and delinquency, not withstanding the role of IQ as an exogenous variable predicting delinquency. In his review of *The Bell Curve*, Duster (1995) notes that a major problem with the book is that the role of IQ shifts from an independent (exogenous) to a dependent (endogenous) variable. "*The Bell Curve* abandons the issue of social outcomes as the dependent variable, and instead introduces the observation that 'even when intervention programs are successful in the short run, they do not alter IQ differences in the long run'" (p. 160). But, as Duster explains, "this is completely contradictory to their fundamental thesis, that IQ is the independent variable producing social outcomes" (p. 160). The argument that Herrnstein and Murray make in the latter portion of their book is that IQ is dependent on heritability and cannot be improved by various interventions. By switching IQ from an independent to dependent variable, the crux of their heritability argument is in jeopardy, thus leaving open the possibility that IQ is not as predetermined by genes as the authors would initially have us believe.

Herrnstein and Murray neglect to examine probable conditions that may antedate low IQ, such as "parental psychopathology, temperamental disturbances, neurological problems, genetic susceptibilities and disadvantageous environmental influences" (Fishbein, 1990, p. 34). Frequently, juvenile delinquents are "multi-vulnerable" populations (Brannigan & Caputo, 1993). For example, if children present from troubled families with fractious parental relations and substance abuse; if they experience frequent neighborhood resettlements as parents separate, become sexually promiscuous before their peers, and are defiant of parental authority before they can realistically live on their own—all of which are common observations to school counsellors—then the link between school trouble, delinquency, and IQ is an interactive process. In such circumstances, delinquency may be a cause of low IQ because students may be removed, discouraged, and demoralized for metascholastic factors, and as they get older frequently absent themselves voluntarily, if surreptitiously, from school attendance. "Discipline" problems are conflated with "scholastic" problems and, unless nonphysical modalities of expression are cultivated, the school breeds children who do not like school, who are poorly motivated, and wither on the vine. They are seen to be delinquent (by teachers) from "their earliest days," and when tested on verbal measures prove themselves to be the numbskulls they were originally conceived to be.

Family conflict resulting in inconsistent support for learning, which is associated with school absence and early school leaving, may itself lower the student's level of literacy. Poor school performance will mirror poor IQ performance, especially in the area of verbal IQ that tests for vocabulary and language use. Again, the causal link between IQ and delinquency may be spurious. Demise of family control structures and support structures may open the adolescent to unrestrained delinquent opportunities (relative to intact families) and fail to provide the family capital (emotional and otherwise) that ensures that the student's native ability is developed to its capacity. The poor IQ performance, especially on verbal and mathematical skills, which are core elements both of school curriculum and intelligence testing, are simply the reverse sides of the same coin: squandered talent. If that inference is warranted, we should not conclude that low IQ warrants apartheid, as Herrnstein and Murray suggest in terms of the "custodial state." Perhaps, rather than spending more on jails as today's conservatives recommend, society should spend more on schools.

References

Agnew, R. (1990). Adolescent resources and delinquency. *Criminology, 28*, 535-566.

Balthazar, M.L., & Cook, R.J. (1984). An analysis of the factors related to the rate of violent crimes committed by incarcerated females. *Journal of Offender Counselling, Services and Rehabilitation, 9*(1-2), 103-118.

Brannigan, A., & Caputo, T. (1993). *Runaways and street kids in Canada in the '90s*. Ottawa: Solicitor General of Canada.

Duster, T. 1995. The bell curve. *Contemporary Sociology, 24* (2), 158-161.

Fishbein, D.H. (1990). Biological perspectives in criminology. *Criminology, 28*(1), 27-72.

Goring, C. (1913). *The English convict: A statistical study*. London: Darling and Son.

Gove, W. (1994). Why do we do what we do? A biopsychosocial theory of human motivation. *Social Forces, 73*, 363-394.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Hirschi, T., & Gottfredson, M. (1983). Age and the explanation of crime. *American Journal of Sociology, 89*(3), 552-584.

Menard, S., & Morse, B. (1984). A structuralist critique of the IQ-delinquency hypothesis: Theory and evidence. *American Journal of Sociology, 89*, 1347-1378.

Naffine, N. (1987). *Female crime: The construction of women in criminology*. Sydney: Allen and Unwin.

Simpson, S.S., & Elis, L. (1995). Doing gender: Sorting out the caste and crime conundrum. *Criminology, 33*(1), 47-81.

Sommers, I., & Baskin, D.R. (1993). The situational context of violent female offending. *Journal of Research in Crime and Delinquency, 30*(2), 136-162.

Townes, B.D., James, J., & Martin, D.C. (1981). Criminal involvement of female offenders: Psychological characteristics among four groups. *Criminology, 18*(4), 471-180.

Ward, D.A., & Tittle, C.R. (1994). IQ and delinquency: A test of two competing explanations. *Journal of Quantitative Criminology, 10*(3), 189-212.

Wilson, J.Q., & Herrnstein, R.J. (1985). *Crime and human nature*. New York: Simon & Schuster.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 322-334*

John R. Kirby
Queen's University

Intelligence and Social Policy

This article addresses the arguments presented in The Bell Curve *in terms of their scientific validity for the context in which they are situated and their value as a basis for social policy in Canada. The article is organized in four sections. In the first I argue that cognitive stratification is less ominous than Herrnstein and Murray indicate, and that the meaning of stratification is dependent on the cultural context. In the second I discuss the nature of intelligence, argue that* The Bell Curve *relies on an impoverished view of it, and suggest that the "intelligence" data they analyze should be seen rather as measuring educational achievement. I describe some of the other limitations in their data analyses in the third section, in particular their underestimate of environmental effects. In the final section, I address two of their conclusions, dealing with removal of support from unwed mothers and their children, and with the shifting of resources from programs for the disadvantaged to programs for the gifted. I argue that Herrnstein and Murray focus too much on symptoms and not enough on causes, and that their analyses are bound to the American context. Contrary to Herrnstein and Murray, I conclude that education for all remains important and that efforts should be continued to understand how intelligence develops and how it can be better nurtured.*

The Bell Curve (Herrnstein & Murray, 1994) is an important book, though flawed and dangerous. It is important because of its scope, its addressing of crucial social issues with an impressive data set, and because it raises the question of "taboo" topics in psychology and education; it has the potential to provoke much thought and debate (see, e.g., the collection of reactions in Jacoby & Glauberman, 1995). It is flawed because of the impoverishment of the concept of intelligence employed, the overreliance on one data set, the failure to recognize contextual effects that may underlie the results presented, and the scant attention given to implications. It is dangerous because its results and conclusions, especially if understood poorly and accepted uncritically, provide ammunition to those attacking whatever is left of the "social safety-net," both in the United States (the focus of *The Bell Curve*) and in our own country, and could be taken as support by those advocating much more radical and frightening steps than those addressed in the book. To address fully the merits and flaws of *The Bell Curve* would require many more than the few pages available to me here. Furthermore, the book could be addressed in terms of either the political philosophy that motivated it or its espoused scientific basis. I opt to treat the book as science and focus on four main areas: the cognitive elite and underclass, the concept of intelligence, the nature of the major data analyses, and the validity of the implications drawn in *The Bell Curve*. My goal is to point toward policy implications for Canada, but this can only be done by

John R. Kirby is a professor of education and teaches educational psychology. His research interests include cognitive psychology and the nature of intelligence.

considering how valid *The Bell Curve*'s conclusions are for the US context in which it is situated.

The Cognitive Elite and the Cognitive Underclass

The first major section in *The Bell Curve* is devoted to "The emergence of a cognitive elite." The basic argument is that over the course of the 20th century more and more intelligent people have been attracted to higher status occupations, those occupations have begun to *require* more and more intelligence, those occupations have become increasingly remunerative (compared with lower-status jobs), and intelligence is associated with job performance, especially for high-level jobs, but even for low-level jobs. As long as the word *intelligence* is broadly (and vaguely) defined, these relationships are not too controversial, though some may be surprised at the speed with which the changes have happened (see the next section for more on the meaning of intelligence).

This argument is quite Darwinian: a characteristic (intelligence) was around in the gene pool, but was not of paramount importance in then-extant environment. One can speculate that physical strength had been important for a long time, replaced by family status. Then the environment changed, presumably by becoming more democratic and complex. Democracy was important, to eliminate brute force as the main determiner of power and to allow social mobility. Complexity was then allowed to exert selection pressure: to master the increasingly complex world more intelligence was required. The trait that had previously been unrelated to success now was so related. Increased intelligence in the work force contributed to the increasing complexity of the environment, increasing the demand for itself. Members of the new cognitive elite increasingly mix, and therefore breed, exclusively with each other. Their children start with environmental advantages in addition to whatever genetic advantages they may have, increasing their chances of staying in the cognitive elite for another generation.

The downside of this story is the development of the cognitive underclass, those individuals whose lower intelligence does not allow them to participate fully in modern society or reap many of its benefits and whose children have little chance of upward mobility. In a classic Darwinian model these individuals fail to thrive, starve, and, most importantly, do not contribute descendants to the future. We humans usually choose not to let these forces take their course; individually and collectively we help the less fortunate around us. Sociobiologists would probably extend Darwinian theory to indicate that by doing so we are helping ensure the survival of our group's genes. Others would say that to do otherwise would be immoral. *The Bell Curve* raises the Malthusian specter of a cognitive underclass growing faster than the elite, surviving or even thriving on the back of the social welfare system. It recalls Aldous Huxley's *Brave New World* in which Alphas, Betas, Gammas, Deltas, and Epsilons were bred to have distinct levels of ability. *The Bell Curve* reminds me of Biggs' (1978) tongue-in-cheek "new Australia," in which there are four classes: managers, producers, mercenaries, and "dolbies," the last of these being welfare recipients prevented from working and rewarded for inactivity and docility.

How real is this cognitive stratification and its consequences? Stratification of some sort seems to be characteristic of modern societies, and the role of intelligence seems more plausible as complexity increases. One has only to open a newspaper to read of the bifurcation of society, one class with increasing wealth and power, whose children enter school with every advantage and aspire to be doctors, lawyers, merchant bankers, and engineers, and the other class with increasing despair and need for social support, whose children aspire to minimum-wage jobs, cleaning up after those who flip the hamburgers. The disappearance of jobs in the middle, and especially at the bottom, only contributes to the seriousness of the situation. *The Bell Curve*'s contribution is to argue that increasingly the stratification will be on the basis of mental ability.

Evolutionary systems are by nature self-regulatory. Mistakes are corrected in time if important. Herrnstein and Murray pay little attention to self-correction within the system other than to argue that the social welfare system subverts self-correction. Consider a few sources of self-correction or instability in the cognitive stratification system. First, if the underclass reaches a certain threshold of despair, there is little to stop them destroying the entire system. Second, if environmental advantages are restricted to the elite class, then it becomes less possible for the more able members of the lower class to move up. The system is based on mobility; as upward mobility becomes impossible, this should lead to less success for the upper class, and the build-up of a more able, and therefore more dangerous, lower class. Third, being in the lower class is bad for your health. *The Bell Curve* reviews some of the evidence that members of the lower class have greater chance of suffering injury or violence; they probably have a greater chance of ill health, being jailed, or of dying young. None of these things contribute to success in breeding. Fourth, environments change, bringing with them the survival value of new traits. Intelligence, broadly conceived, must be polygenetic and must change in nature (though perhaps only slightly) as the environment does. To the extent that the elite isolates itself from the rest of the world, it sows the seeds of its own demise. Fifth and most importantly, one of our human traits is morality, entailing care for others. Whether seen in an evolutionary light or not, morality will be a force acting to attenuate the ill effects of stratification.

It is important to consider the *size* of the cognitive class that Herrnstein and Murray consider to be a problem. The second part of the book, "Cognitive classes and social behavior," describes the association between intelligence and a variety of social problems (more on this when I consider their data analyses in detail). The vast majority of the social problems (including poverty, school failure, welfare dependence, and crime) are evidenced by the lowest 10% or so of the cognitive distribution (after controlling for social class and other variables). Of this group, of course, only some show social problems, ranging from small percentages up to 40%, depending on the characteristic in question. Furthermore, the vast majority of the population is not really contributing to the relationship between intelligence and social problems (the graphs flatten out above –1 standard deviation). Although 10% of the US population is a large number in absolute terms, the portion of it demonstrating social pathology is not about to threaten the stability of the country. Perhaps more importantly,

these results may be the consequence of the US' peculiar history (especially with respect to race), and different social policies (e.g., better support for the poor) may alter the relationships.

The existence of cognitive stratification does not establish the *meaning* of that stratification. That different groups of people carry out different jobs and are paid differently does not automatically cast scorn on the "lower" group or invalidate them as members of society. The US may represent a relative extreme in the treatment of the less fortunate; some other societies, for instance much of Western Europe and Australia, seem to treat their less fortunate citizens with greater dignity through government action; other societies, for instance those of east Asia, may rely more on families to care for the less fortunate. Through taxation and welfare, governments can equalize incomes, regardless of cognitive status. Being on "welfare" does not carry the same social stigma in other countries as it does in the US.

We should not lose sight of the role of education in dealing with the effects of stratification, not just education designed to raise intelligence *per se* (see Implications for Social Policy section), but also education designed to impart valued job skills. Even if we accept *The Bell Curve*'s arguments that intelligence cannot be raised and that job performance is correlated with intelligence, all is not lost. Education can contribute by helping individuals maximize their achievements and by helping them understand their society better. Education becomes even more critical as industrial or less-skilled jobs disappear if we want to maximize the number of persons participating in the coming "information age" society (Reich, 1995). In many cases it is also possible to change the jobs, making them less demanding of rarer skills.

Whether one likes it or not, some form of stratification seems to be occurring, especially in the US, and the chances are great that intelligence is associated with it. In the absence of comparable Canadian data, one can only speculate that the same trends exist here, perhaps at an earlier stage of development and, one hopes, with greater attenuation through the social system.

The Nature of Intelligence

The preceding section assumes a broadly based definition of the term *intelligence*. Discussion becomes more controversial once one attempts to measure intelligence, especially if one posits that intelligence is (mainly) one thing (*g* or general intelligence). The debate becomes quite polemical once the heritability of intelligence is addressed, and worse still if racial or ethnic differences are proposed. Although it is absolutely necessary to measure intelligence if it is to be studied, the other proposals are more open to debate. Herrnstein and Murray argue that intelligence has become a taboo topic (p. 1), but it is more accurate to say that the taboo is confined to the latter proposals, those concerning heritability and group differences. Over the last three decades those promoting the importance of intelligence have claimed to have the facts on their side, whereas the opposition has focused more on the moral repugnance of these hypotheses. This has given rise to another myth in the minds of laypersons (including many in education and psychology) that this is another instance of science versus emotion. Herrnstein and Murray seem keen to support this myth, taking the side of the "good" science and claiming with it the force of science in its other manifestations. One can observe that the political

right wing, especially in the US, is doing likewise at the moment, aiming to cut welfare, though they are not as likely to do so when science opposes the teaching of creationism or supports sex education, birth control, or gun control. There are, however, serious reasons to doubt the scientific validity of the proposals concerning *g*, heritability, and group differences. Science is by no means wholly on the side of *The Bell Curve*.

General Intelligence

There has been debate for most of this century whether one factor provides a sufficient model of intelligence (Kirby, 1980). Although much of the debate has focused on the data and data analyses, there has been increasing interest in theoretical interpretations. On the data and data analysis side my impression is that an official tie has been (or should be) declared. In spite of the wonders of confirmatory factor analysis, the problem remains that the factors that emerge depend on what measures have been given to what subjects, the algorithm used in analysis, and the judgments of those analyzing the data. Empirically there is considerable support for a hierarchical model, with *g* at the top and increasingly narrow, but somewhat correlated factors at successively lower levels.

With respect to theory the field is also divided. For most of this century practice has accepted the notion of general intelligence, perhaps divided into verbal and spatial components; thus the infamous but ubiquitous IQ scores. Spearman's (1904) initial characterization of *g* as "mental energy" did not help much. Although a single factor can be obtained in any analysis, it is not clear whether this mathematical entity has any psychological reality, though its convenience cannot be denied. Cognitive psychology entered the fray on the side of *multifactorial* interpretations of intelligence, arguing that the vast array of cognitive processes cannot be understood one-dimensionally. Sternberg (1988) proposed that intelligence consists of metacomponents, performance components, and knowledge acquisition components and argued that socio-cultural context and the individual's learning experience must be taken into account. Gardner (1983) has suggested seven forms of intelligence: verbal, musical, logical-mathematical, spatial, bodily-kinesthetic, and two forms related to one's intrapersonal and interpersonal skills. Das, Kirby, and Jarman (1979; Das, Naglieri, & Kirby, 1994; Kirby & Das, 1990) have argued for the importance of four factors, Planning, Attention, Simultaneous, and Successive processing (the PASS theory). Cognitive research has recently provided some support for *g* too, in the form of speed of mental processing (pp. 284-286). The multifactorial theories tend to be more optimistic about the possibility of increasing intelligence because they deal in smaller, easier-to-handle constructs; it is easier to see how "planning" might be trained than "speed of mental processing."

My point is that there is neither empirical nor theoretical agreement among scholars in the field on the status of *g*. *g* is not dead, but neither does it rule. For many scholars *g* is the question, not the answer. Furthermore, the multifactorial cognitive models offer means to understand how intelligence develops and how its development may be encouraged.

Measuring g

The vast majority of the data presented in *The Bell Curve* comes from one study, albeit a large and important one. The National Longitudinal Survey of Youth (NLSY) has collected a staggering amount of data from over 12,000 subjects, beginning in 1979 when the subjects were aged 14-22, and continuing to the present (pp. 118-120). The measure of intelligence employed was a composite of four subtests of the Armed Forces Qualification Test, word knowledge, paragraph comprehension, arithmetic reasoning, and mathematics knowledge, all administered in the second year of the study when the subjects were 15-23 (Appendix 2, for further details). This measure is then used to predict later social behavior in the 1980-1990 period.

Four points should be made about these research methods. First, no serious scientist would propose these methods a priori to answer the questions addressed in *The Bell Curve*; Herrnstein and Murray were fortunate to have access to such a large and comprehensive database, but had to live with its measures and times of testing. Second, the measures do not derive from any known *theory* of intelligence; they are merely measures that seem to reflect important outcomes (much the same can be said of most ability tests: they represent the triumph of measurement over theory; see Kirby, in press, for more on this). Third, although the subtests may be highly "*g*-loaded" (i.e., highly correlated with more widely accepted measures of *g*), they are clearly measures of school-learned knowledge and skill; thus any factor that has contributed to success in schooling is implicated in any variable these scores predict. Fourth, the measures were administered in adolescence or early adulthood after many other factors have had their influence: in no sense is the IQ score used in *The Bell Curve* a plausible measure of "raw" talent. I believe these last two points combine to provide *The Bell Curve*'s Achilles heel.

But that does not mean that the data are nonsense. Although I have not seen the tests, they seem to provide a good all-round estimate of cognitive performance where cognitive performance is explicitly understood as the unanalyzed product of heredity and environment. As such they are probably quite useful for predicting subsequent behavior, but not for understanding the fundamental causes of that behavior or the means by which intelligence develops. I suggest that the IQ variable used in *The Bell Curve* be thought of instead as "educational achievement"; this does not invalidate many of the relationships that Herrnstein and Murray present in Chapters 5-16, but it makes some of them less profound (Is it any surprise that students of lesser educational achievement do worse in school?), and radically alters the implications. I expand on this in the section Conclusions for Social Policy.

Genetics and Heritability

I am astounded that few people distinguish between genetics and inheritance. To put it simply, you may inherit your parents' silverware, but the mechanism is not genetic. The same can happen with intelligence. One may share a mental characteristic with one's relatives without the mechanism being genetic. It is difficult to determine the extent to which such a characteristic is determined genetically as opposed to environmentally, and in many cases the question does not make any sense. The problem remains in the rare instances of identical twins reared apart, because "environment" starts at conception, a point at

which few twins are separated. Even that may not be soon enough, because it is possible that egg or spermatozoon had been damaged in some as yet unknown manner. Current behavior, even on a highly *g*-loaded ability test, is a function of genetics, environment, and their interaction. Current estimates of the heritability of intelligence (the amount of variance in IQ due supposedly to genetic factors, largely derived from identical twin studies) are contaminated, and undoubtedly inflated, by two factors: the shared early environment (even if only prenatal and perinatal) and the fact that adopted children are not randomly assigned to environments. Furthermore, any estimate of heritability is limited to the context in which the estimate was obtained; if you change the environment, for instance by making radical changes to children's early environments, the estimates may no longer be valid. I accept *The Bell Curve*'s estimate of .6 to .8 for the heritability of intelligence if these limitations are understood, but I do not see that this estimate constrains what *might* happen in the future. Heritability may be more a result of the way things are than a cause of it.

Racial and Ethnic Differences

It is not at all clear to me why Herrnstein and Murray spend so much time and effort presenting their analyses of racial differences in IQ and social behavior. Presumably in a democracy, especially one with a Bill of Rights (even one whose signatories included slave-owners), one is not going to do anything as a consequence of racial differences other than attempt to attenuate their worse effects. Herrnstein and Murray often make the point that individuals should be treated on their own merits, not as a function of their race. So why do they devote so much space to racial differences? I fear that there is a considerable market for such information, and that even though the purveyors of the information insist that it should not be used as the basis for decisions, it will be.

I believe that there are good scientific reasons for discounting racial and ethnic group differences beyond any moral reasons that may be offered. First, race is not a very satisfactory biological variable. Those who study race as a variable usually employ it as a measure of some innate difference. Yet it is not clear how race is determined or how it is thought to act as a biological variable. If you have 15 "white" ancestors and one "black," but have a dark skin color, you are more likely to be treated as "black" than someone who has 15 "black" ancestors and one "white" but has a lighter skin color. I presume that it is not skin color per se that is thought to influence cognitive abilities, but rather a complex of genes that are argued to be associated with it. It seems unlikely that the gene or genes responsible for skin color are also responsible for mental ability, so the effect (if there is one) must be due to genes that are merely associated (i.e., correlated) with them. Given the amount of mixing that has taken place over the centuries, particularly in the US, it seems unlikely that many gene associations would remain distinct; in any case, there is no necessity for them to remain distinct in the future. Perhaps race should be seen more as a self-report variable like self-concept; instead of a pure biological, causal factor, it is more the consequence or accretion of an array of causal factors that are still poorly understood. Herrnstein and Murray acknowledge that they employ race in this way, as one's self-classification (p. 271), but then go on to "comment on cognitive differences among races as they might derive from genetic dif-

ferences" (p. 272). Although they admit at the outset that there are "more questions than answers" (p. 272), one wonders if their readers keep this qualification in mind as they read the book.

Second, like most self-report variables, race is not very useful in causal analyses, as you cannot randomly assign subjects to different levels of it. No matter how sophisticated the statistical analysis, race is something the subjects come in the door with. The more you believe in racial differences, the less you should be able to believe, for instance, that the same SES score means the same thing for members of different races.

Third is the notion of bias. Herrnstein and Murray do a good job of discussing the evidence on various forms of bias in mental testing. They concur with the literature (Jensen, 1980) that the commonly used mental tests are not biased either at the item level or in terms of what they predict for blacks and whites in the US. They admit that there is a form of the bias hypothesis that is not directly subject to empirical investigation: "the tests may be biased against disadvantaged groups, but the traces of bias are invisible because the bias permeates all areas of the group's performance" (p. 285). Herrnstein and Murray term this the "background radiation" effect and discount it, but I find it quite plausible. Their argument against it is based on the finding that blacks do *not* do worse than whites on some measures; these measures (Forward Digit Span) are not related strongly to the core construct of intelligence employed in *The Bell Curve*, so I do not see this as a great difficulty.

For these reasons, I choose to ignore *The Bell Curve*'s Chapters 13-16 on racial and ethnic differences.

The Bell Curve's Data Analyses

The core of *The Bell Curve* is in Chapters 5-12, which present the effects of intelligence ("educational achievement," as I argue above) on a variety of social behaviors. The basic question addressed is: What effect does intelligence have on behavior after controlling for socioeconomic status (SES)? Regression analysis is used throughout, and a reasonable amount of detail is provided in Appendix 4. The results indicate that intelligence (as measured in the NLSY) is more strongly associated (negatively) with a range of social problem behaviors than is SES (again, as measured in the NLSY).

Multiple regression is a powerful and sophisticated technique, but no statistical technique is any better than the data fed into it. As techniques become more sophisticated, the nature of the assumptions on which they rest and the quality of the reasoning process become more obscure, especially to the lay reader. The regression analyses in *The Bell Curve* suffer from three classic limitations.

First, statistically significant results do not necessarily mean practically important results. Herrnstein and Murray themselves admit that many of their results are weak, intelligence accounting for less than 20% (and often much less) of the variance in behavior. I fear that many readers will miss this point. Such results may be important for theory, but offer little to practice.

Second, the constructs of interest may be poorly measured by the actual variables. I argue above that *The Bell Curve*'s intelligence measure might be better termed educational achievement. Their SES measure is probably typical of the literature, being the average of total net family income, mother's educa-

tion, father's education, and an index of occupational status (p. 574), but one can question whether these measures truly tap what is meant by "environmental effects." It seems obvious that the quality of the home life, especially as it relates to cognitive behavior, is most likely to be the causal agent and that it is at best poorly measured by the SES index. Thus their SES measure may only address a small proportion of environmental effects.

Third, a significant regression effect between Predictor A and Outcome B can give the impression that A is a cause of B. Regression analysis makes use of correlations, and all the cautions appropriate to correlational analyses need to be applied to regression analyses, especially "Correlation does not imply causation." When predictors such as *The Bell Curve*'s IQ measure are employed, one must be quite suspicious whether they are predicting successfully because they are associated with other variables that are actually the causal agents. For example, educational achievement in late adolescence may be the complex result of early environment, language stimulation, and education above and beyond genetic effects. A significant regression effect could be the result of these factors rather than "raw talent," and would not be attenuated by the SES variable that does not measure such things. In technical terms, one wonders if the regression models are underspecified, that is, whether they contain the "right" predictors.

Fourth, regression results, like any others, apply only to the context in which they were derived. This seems particularly relevant when the constructs and variables are laden with social import and when many of them (e.g., being in "poverty") are a function of laws and social policies. To illustrate: if the poverty line is raised or lowered, the number of poor people is altered, when in fact no one's income has changed. A given level of income may have very different consequences depending on where one lives and the generosity of friends and family. To be more extreme, if every citizen received a guaranteed annual income that was above the poverty line, none would be poor at all. The results shown in *The Bell Curve* are for the US, and are therefore dependent on their specific history. One of the disappointments of *The Bell Curve* is that the authors did not regard other countries as "experiments" that could shed light on American problems (it is also possible that there is a lack of data). One can wonder what effect a history of racism, violence, militarism, robber baron capitalism, and crime-linked drug abuse has on the relationships among intelligence, SES, and social behavior, but one cannot do regression analyses on the NLSY data to find out.

Implications for Social Policy

In the final section of the book we are told that attempts to raise intelligence have failed, that American education is doing a better job with the less able than with the more able, and that affirmative action is a cancer destroying American institutions. Herrnstein and Murray have some suggestions: restore many social functions to the neighborhood, simplify the rules by which society operates, reorient education, modify affirmative action policies, discourage low-income/ability women from having children, and modify policies so that immigration is based more on competence and less on humanitarian or family criteria (these are presented in Chapters 17-22).

In general the recommendations conform to the current right-wing American agenda, aiming to eliminate federal government involvement, regulation, and social welfare programs without much thought to why they were initiated. They do not strike me as profound or insightful. More disturbingly, they concentrate on the symptoms of problems rather than their causes. Many of the recommendations are only loosely related to the extensive analyses that precede them.

I focus on two recommendations having particular salience for educators, the removal of legal rights to support from unwed mothers and their children, and the shift of resources from disadvantaged to gifted programs. Herrnstein and Murray's obsession with marriage typifies their focus on symptoms, not causes. Like many commentators in the US at the moment, they seem convinced that unwed mothers are luxuriating on welfare and are motivated to have children by the promise of welfare support. Their argument is not supported with much evidence, ignores much relevant thinking, and lacks plausibility. Has anyone ever surveyed unwed mothers to find out why they have children? This seems like an important first step. My guess is that the reasons are quite varied, ranging from the middle-class woman with adequate resources who decides she wants a baby but not a husband, to other women, probably much younger, who desperately need to be loved, lack the self-esteem to say no, lack knowledge about birth control or lack the mental ability to use that knowledge, and have no access to abortion. There seem to be many ways of tackling these myriad factors instead of removing social support. Removing that support seems to me to guarantee that even more children will be raised in poverty by the least adequate mothers with less and less assistance. If there is a cognitive underclass, surely this is its breeding ground. Furthermore, the identical treatment of all unwed mothers and their children begs the question of what the causal factors are. I can only assume that Herrnstein and Murray do not think that having a marriage license tucked away upstairs will prevent poor parenting or that having a violent and abusive spouse in the home is preferable to none at all. My point is that poor parenting is the problem, not the absence of marriage ceremonies. Let's find out what the causes of the problem are and eliminate them.

The recommendation that resources be shifted from disadvantaged to gifted programs is based on two arguments: that efforts to raise cognitive ability have failed (Chapter 17), and that American education has been "dumbed down," that is, made easy so that the less able can pass (Chapter 18). Even if these arguments were true, shifting resources to gifted programs would not follow. First, many of the programs (e.g., Head Start) reviewed in *The Bell Curve* either were not designed to raise intelligence, or should not have been, but rather aimed to increase achievement either in school or the world (see Zigler's comments in Azar, 1995). Second, that particular programs have failed does not imply that the task is impossible, just that we have not yet found out how to do it. It seems plausible that the key to raising intelligence is the quality of parent-child interaction in the early years, a factor requiring parent education in addition to special child education. Until all such factors have been addressed in interventions, we should not conclude it cannot be done. We may then conclude that we cannot afford it, but that is a different issue. Third, even

if American education has been pitched toward the lower end of the ability spectrum, there is little evidence that it has been very successful—in fact much of *The Bell Curve* can be seen as an indictment of that lack of success. Should the goal of raising performance of low and middle achievers be abandoned because it has not been done well? Fourth, much of *The Bell Curve* seems to argue that the cognitive elite are getting more and more, both in terms of education and employment; in what sense has education failed these individuals? Perhaps the point is that we *could* have more such individuals, but *The Bell Curve* argues that virtually all of the cognitive elite can already be located in the elite professions or at the upper levels of business (p. 60); there seems to be a contradiction here. Fifth, why must programs for the gifted come at the expense of programs for the less able?

Conclusions

Although I suggest that there are many weaknesses in *The Bell Curve*'s arguments, there is merit in some of the issues raised and in some of the conclusions reached. Canada does have something to learn here, though not always what *The Bell Curve* would intend. I confine my conclusions to two areas, the importance of education and the effects of the American context on the analyses presented, and then make final comments about politics and intelligence.

Education. Contrary to *The Bell Curve*, my reading of their analyses only strengthens my belief that education is important for all, not only those at the top. The main dependent variable that they analyze, termed by them intelligence, can be argued to represent instead educational achievement. No doubt intelligence affects educational achievement. Even if intelligence controls 99% of the variability in achievement and is .99 heritable due to genetic factors, the purpose of education is to address whatever is left over. As many have argued, it makes more sense to address what we can change rather than what we cannot. To use an argument used elsewhere in *The Bell Curve*, very small effects, even .1 SD per generation, will accumulate and compound in time. If society is changing and lower-skill positions are being eliminated, education in the form of skill training will be more and more important. Even if there are not enough jobs to go around, it will still be important to have a literate and informed populace. Literacy is within the reach of all but the least able, but we are not yet close to that level of success. And we should not give up on attempts to understand how intelligence develops and whether it can be nurtured more successfully. Regardless of how the Americans decide to change their education system, we should devote ourselves to maintaining and strengthening Canadian education systems.

This does not mean that all *The Bell Curve*'s contentions and proposals regarding education are wrong. It is worthwhile to consider whether literacy and numeracy standards are sufficiently high, whether the most able students are adequately supported and challenged, and whether parents should have more control over what school their children attend. But these proposals are already part of the public debate and are being acted on in various provinces. *The Bell Curve*'s contribution is to emphasize the importance of intelligence, both as raw talent nurtured by schools and homes, and as the outcome educational achievement, which their data suggest is important in preventing a host of social ills.

Social context. We must remember the context in which *The Bell Curve*'s data and results exist. Would the same results occur in a society where there was less absolute poverty, where the poor were treated with greater dignity and had better access to health and other services, where there was less violence, fewer guns, less of a necessary link between drug usage and crime, more access to birth control and abortion? We do not know, and *The Bell Curve* cannot tell us. At the very least it seems reasonable to attempt to maximize the quality of the social context, to minimize the deleterious effect that many variables could have. There is simply no evidence that *decreasing* the quality of the social context will have any benefit, except perhaps for those whose taxes will go down, and then only in the short term.

Politics. It would be hard to conclude without touching on the politics that lurk just below the surface of the debates triggered by *The Bell Curve*. *The Bell Curve* is science addressing social issues, and that is good. But to what extent is it relatively dispassionate scientists examining the data and reaching conclusions that just happen to fit a particular political model, rather than being individuals of a particular political persuasion searching for data, results, and arguments that match their politics? As I suggest, *The Bell Curve* is consistent with the right-wing agenda in the US these days, and it has evoked the predictable left-wing objections (Jacoby & Glauberman, 1995). There is a tendency toward defending one's "wing," regardless of the merits of its case. This is a sure recipe for false dichotomies and sterile debates to which education is no stranger. If truth lies between these extremes, we will have to listen to the arguments of both sides and judge each point on its merits. To believe in intelligence is not necessarily to accept strong genetic arguments, and to believe in education is not necessarily to oppose scientific enquiry. None of us is unbiased on these issues, but that does not mean that we cannot consider each argument and proposal fairly and intelligently.

Intelligence. In the end, what relevance does intelligence have for social policy? Seen as educational achievement, *The Bell Curve*'s evidence is that it is quite important, given the many limitations discussed above. Greater achievement could well reduce many of the social problems that plague our society, though we cannot be certain of that from *The Bell Curve*'s analyses. *The Bell Curve*'s data do not address intelligence as raw talent (i.e., the cognitive ability with which children enter school), but it is reasonable given much other research to assume that raw talent is one of the major contributors to educational achievement. If we wish to increase raw talent, there is really no option in a democracy to improving the quality of early environment and education, regardless of the degree to which raw talent is a function of genetics or early environment.

Acknowledgments

Much of this article was planned while I was on sabbatical at the Universities of Newcastle and Western Sydney in Australia. I thank those who discussed *The Bell Curve* with me, especially Robert Cantwell, Trevor Cairney, Philip Moore, and Neville Schofield. I also thank Robert Wilson for commenting on a draft of this article and for our continuing discussions of regression analysis.

References

Azar, B. (1995). DNA-environmental mix forms intellectual fate. *APA Monitor, 26*(1), 24.

Biggs, J. (1978, January 9-14). The new Australia. *National Times*.

Das, J.P., Kirby, J.R., & Jarman, R.F. (1979). *Simultaneous and successive cognitive processes*. New York: Academic Press.

Das, J.P., Naglieri, J.A., & Kirby, J.R. (1994). *The assessment of cognitive processes: The PASS theory of intelligence*. Boston, MA: Allyn and Bacon.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Jacoby, R., & Glauberman, N. (Eds.). (1995). *The bell curve debate: History, documents, opinions*. New York: Times Books.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Kirby, J.R. (1980). Individual differences and cognitive processes: Instructional application and methodological difficulties. In J.R. Kirby & J.B. Biggs (Eds.), *Cognition, development, and instruction* (pp. 119-143). New York: Academic Press.

Kirby, J.R. (in press). Test review: The Woodcock Language Proficiency Battery. *Journal of Psychoeducational Assessment*.

Kirby, J.R., & Das, J.P. (1990). A cognitive approach to intelligence: Attention, coding and planning. *Canadian Psychology, 31*, 320-331.

Reich, R.B. (1995, April 2). Drowning in the second wave. *New York Times*, p. 15.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-209.

Sternberg, R.J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Penguin.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 335-341*

David W. Livingstone
Ontario Institute for Studies in Education

For Whom *The Bell Curve* Toils

Intellectuals and Monuments

It is evident to even the most casual student of past civilizations that the ruling classes typically have had considerable material and intellectual resources devoted to the construction of monuments promoting their own superiority over the ordinary folk. Our own times may seem to us to be more complicated, both in terms of the relations between groups with economic or political power and diverse intellectuals, and in terms of the artifacts we could leave behind. For most contemporary intellectuals most of the time, the relationship between their particular work and the interests of any supposed ruling class remains quite obscure. But even in advanced industrial market societies such relations have become more evident in times of social crisis when the old social contract needs to be reconstructed.

As adept biographers have documented since the mid-1800s, the private mindsets of corporate elites have quite consistently presumed the inherent superiority both of their own capabilities and of market dynamics over government intervention. But in modern liberal democracies, elites have frequently reconciled themselves to more progressive institutional forms negotiated with the lower classes. When the old social order breaks down, both new economic deals and alternative political ideologies are needed to build a new one. It is in these historical moments that the alignments between intellectuals' knowledge and social groups' powers is at its most apparent. Invariably in such social crises, arguments in favor of the inherent superiority of a ruling group in terms of some discernible feature have been promoted by affiliated intellectuals. The bell curve, or "normal" distribution of IQ test scores, has been used for generations to legitimize the grading and selection of students. It may ultimately be found to be one of the key justifying artifacts of hierarchically organized industrial capitalist societies. Herrnstein and Murray's *The Bell Curve* is more likely to be remembered as an early right-wing, corporate-sponsored bluff in the current renegotiation of our social order. The extraordinary attention given to this book is probably more closely related to the resonance of its arguments with the private world views and neoconservative political agendas of the corporate controllers of the mass media than with any scholarly merits of the book itself.

In a democratic society anyone has the right to argue for any viewpoint. But it is only fair to expect advocates to identify what vantage point they are arguing from, as well as what they are arguing for. In a class society, totally independent critical inquiry is an illusion. The best we can hope for in social

David Livingstone is a professor and chair of the Sociology Department. His publications include *Class, Ideologies and Educational Futures, Social Crisis and Schooling, Critical Pedagogy and Cultural Power, Stacking the Deck,* and *The Education-Jobs Gap.*

research is frank declaration of interpretive vantage points and political alignments, the conduct of replicable empirical studies, and the honest presentation of findings and conclusions. Charles Murray is a Bradley fellow, as noted on the dustjacket of this book. The Bradley Foundation is one of largest sponsors of conservative intellectual work in the United States. He and Herrnstein (Herrnstein & Murray, 1994) fail to note that most of the research on which the central claims of *The Bell Curve* are based—including that of Arthur Jensen, William Shockley, Philippe Rushton, Richard Lynn, and Thomas Bouchard—has been underwritten by the Pioneer Fund, a small endowment fund established by a Massachusetts textile heir to advocate hereditarianism (Lane, 1995). Herrnstein and Murray are advocates for hereditarianism masquerading as independent scientific researchers. By the same token, most of their many critics to date also pose as independent intellectuals while advocating environmentalist perspectives and implicitly aligning themselves with existing liberal institutions. This particular commentary comes from a professor at a public university whose research on issues of social inequality has been funded by government-sponsored agencies through peer-reviewed competitions. My political sympathies are definitely with subordinated groups, particularly lower class people. I have no fixed position on the nature-nurture debate, but remain skeptical of arguments based on either the inherent superiority of designated groups—such as that of Herrnstein and Murray—or the presumed adequacy of established institutional forms—such as some of the counterarguments of their legion of critics (Lane, 1995).

IQ Scores and Jump Shots

Herrnstein and Murray conveniently lay out most of their underlying assumptions/conclusions early in the book (pp. 22-23) prior to launching into their linear regression-based ransacking of the National Longitudinal Survey of Youth. In a nutshell, there is a general factor of cognitive ability; it is most accurately measured by IQ tests; IQ scores match what people mean by "smart" in ordinary language; IQ scores are stable through life and unbiased; and cognitive ability is substantially heritable. The final key assumption stated later is, of course, that the dominant economic class is increasingly constituted by the most intelligent people.

The entire argument, therefore, hinges on the adequacy of IQ tests as a summary measure of intelligence. I have personally distrusted IQ test results ever since my first term in high school. I took an IQ test that seemed so offensive in its standardized multiple choices that I was provoked to spend about half an hour writing an angry critical essay, which I duly enclosed and for which I received zero IQ points. One of my core academic teachers, who was a true believer in the validity of IQ as a measure of intelligence, soon labeled me as an "overachiever" who didn't have the real talent of some of my "gifted" classmates. I confirmed this assessment from an unimpeachable source, a new girlfriend who went to another school but whose sister was a teacher at mine and who witnessed these characterizations in staff meetings around student awards decisions. After experiencing the consequences of this labelling for a few months, I declared my criticisms of the test and this teacher to the school principal. My academic image seemed to improve almost immediately. As a white, middle-class boy with high self-esteem, I sensed that this

low intelligence label was both wrong and dangerous to my future. In fact, because I'd spent more time shooting jump shots than doing homework, I was much more proficient as a basketball player than a student at that point. A few years later, my dream of becoming a professional basketball player had been shattered, but I could score on any IQ test (Esterbrook, 1995).

More generally, the notion that our cognitive abilities can be reflected adequately in a single summary measure is extremely suspect. For example, the Harvard psychologist Howard Gardner has identified at least seven distinct intelligences, including linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, and intra- and interpersonal computational capacities. Herrnstein and Murray dismiss such views as "radical" because of an absence of statistical verification (p. 19). Even if these various intelligences tend to be modestly correlated, this does not warrant reducing them to a single number. It would be just as reasonable to treat measures of these abilities additively—but I doubt that my classmates would have been any more impressed by my "genius" if my points-per-game average had been added to my IQ score.

The basic problem with IQ tests, however, is their cultural bias. IQ tests have usually been designed by upper-middle-class experts, using middle-class linguistic codes and cultural knowledge content. These tests are administered by middle-class teachers mainly to aid in the selection of the middle-class workers of the next generation. There are systemic barriers to the effective performance by immigrant and lower-class kids on such tests, whatever their actual intelligence. If intelligence is assumed to be what these IQ tests measure, then lower-class kids will always be found on average to be less smart. In the most extreme case, to be illiterate is to be presumed stupid. But consider the comment of a rural Vermont illiterate worker:

> It don't bother me. They can't buffalo or fool me on anything. As far as doing any kind of work, I can do any kind of work that anybody can. I'm far enough ahead on stuff, outside of reading, it's quite a job for anyone to trick me on anything. There ain't a piece of equipment yet that I can't run, or I can't tear down and put back together. (Cole, 1976, p. 66)

This appears to be one "smart" guy. Popular education programs with poor illiterate people around the globe have found that they can commonly become literate in extraordinarily short periods when working with sympathetic instructors and using curricular materials that reflect their own experience.[1] Using Herrnstein and Murray's criteria, these people would simply be written off.

IQ scores are far less immutable than Herrnstein and Murray assert. There have been average increases in IQ scores in some countries during the post-WWII period equal to the current 15-point difference between US blacks and whites. Poor black kids adopted into affluent white homes have shown substantial increases (Gould, 1995). Herrnstein and Murray do recognize an upward drift in IQ scores generally and also concede that the IQ gap between whites and blacks has narrowed by about three points in the past few decades, probably because of blacks' improved material circumstances. But they argue tautologically (Chapter 13 and Appendix 5) that this convergence is likely to cease because of higher fertility among lower-IQ blacks who are presumed to hand on lower IQs to their offspring. Whatever may happen in the future, such

historical changes reveal IQ tests to be socially constructed devices to measure normatively approved cultural knowledge rather than any inherent universal cognitive ability (for a revealing historical account, see Rose, 1979).

Perhaps the most extraordinary aspect of *both* hereditarian and environmentalist approaches to intelligence based on IQ scores is a general failure to conduct empirical studies that measure *parental* IQs. The most elementary test for the heritability of IQ score would be the intercorrelations of mother's and father's IQ scores with that of the child. *The Bell Curve* is filled with speculative calculations on the extent of heritability based on the assumption that it is somewhere between 40% and 80%. Although generally high correlations between the IQ scores of identical twins raised separately are suggestive of high heritability of some cognitive abilities, only parent-child IQ score correlations could address the question of heritability directly. But the perspective of parental genetic transmissions immediately raises mitigating conditions such as other factors besides intelligence in mating affinities, as well as random variation and regression toward the mean in the combination of parental traits. An even more profound limitation to this whole form of debate is the fact that advances in the neurosciences and molecular biology in recent decades have discovered complex interactions between brain development and environmental experience. So either to assign ballpark estimates to genetic factors as Herrnstein and Murray do, or to dismiss genetic transmission as many of their critics do, is to display one's ignorance of the existence of this feedback loop (Duster, 1995).

Cognitive Elites and Underemployment

Herrnstein and Murray profess great concern about the partitioning of the smart and the stupid, with the growth of a cognitive elite of high-IQ professions at the top and low-IQ unemployed and welfare cases at the bottom. Certainly the economic slump that began in the advanced industrial market economies around 1970 and the subsequent tax and services cuts by neoconservative regimes have led to class polarization, with the rich getting richer and the poor underclass burgeoning. But there is little evidence for the sort of partitioning on the basis of intelligence that these authors suppose. In fact most of their arguments about employment relations are vitiated by their admitted failure to cover "large macroeconomic forces" (p. 157). The long-term decline of agricultural self-employment and the post-WWII expansion of professional-managerial jobs, along with the "credentialism" associated with the massive expansion of postsecondary schooling, are contextual effects that go a long way to explain any increasing concentration of more highly educated people in professional jobs.

Even the simplest estimates of intergenerational reproduction of occupational classes indicate that Herrnstein and Murray's views on cognitive partitioning across classes are quite exaggerated. Table 1 presents recent estimates of occupational class reproduction for the Ontario male labor force based on a series of surveys that I conducted and using widely accepted Pineo-Porter measures. Although these surveys contain no direct measures of IQ, they do suggest two important points about class formation. First, the intergenerational reproduction of corporate executive or professional statuses is achieved by only a minority. Second, there is clearly intergenerational regression toward

Table 1
Intergenerational Reproduction of Occupational Class, Ontario Male Labor Force, 1988-1994

Respondent's Class	*Occupational Class Matches Father's Age Cohort*				*Column*
	18-34	*35-44*	*45-54*	*55+*	*Total*
	%	*%*	*%*	*%*	*%*
Corporate executive*	—	27	29	43	<1
Professional/High Manager	39	24	33	22	14
Semi-professional/Supervisor	35	20	19	26	24
Skilled/Semi-skilled	58	53	64	65	44
Unskilled	38	29	30	23	17
N	378	364	256	225	1,223

*Including professional/high manager designations on Pineo-Porter Scale.
Source: Livingstone, Hart, and Davie (1995).

the middle of the class structure from both the top and the bottom. No *direct* evidence is presented by Herrnstein and Murray to indicate that the reproduction of intelligence within classes is any more pronounced.

But even more damaging for this book's central thesis of an increasingly class-based distribution of intelligence is the existence of substantial and growing *underemployment* (Livingstone, in press). In spite of credential inflation, a large proportion of the work force now has more schooling than their jobs formally require. Many highly educated people can find no jobs at all; for example, over 30% of the users of food banks in Toronto last year had postsecondary education (unpublished survey, Daily Bread Food Bank, May, 1994). So, even if one accepts Herrnstein and Murray's assumptions linking advanced schooling and intelligence, the number of bright people clearly remains much greater than the jobs for which high intelligence is needed—quite the contrary to their assertions (p. 27).

Custodial Darwinism?

After over 500 pages of speculative arguments and analyses of this order, Herrnstein and Murray put forward their policy proposals. The central thrust is captured in two chilling sentences: "Cognitive partitioning will continue. It cannot be stopped, because the forces driving it cannot be stopped" (p. 551). In other words, those who make low scores on IQ tests should be content with menial positions in the future because it is inevitable that those with high IQs will rule the world. Although the cognitive elite rules, most folks will just get on with their lives as individuals in a "simplified" state where group differences are "insignificant." Such differences are insignificant, that is, unless one happens to be a recent immigrant or a black single welfare mother, in which case they are seen as more likely to be stupid and therefore subject to "social-engineering solutions" (pp. 549-550). Such solutions remain unspecified except for prior allusion to a custodial state with "high-tech Indian reservations" (p. 526) for stupid deviants. My own recent surveys have, rather, found visible minority immigrants to be more likely to be underemployed, with relatively

higher levels of schooling than white Canadians, whereas women generally are more underemployed than men (Livingstone et al., 1995). But never mind such aberrations. The essential law is superior survival for those with the fittest IQs, who just happen to be preponderant in the white upper middle class. Herrnstein and Murray serve up a bizarre combination of the crudest social Darwinism without even a simple explanatory mechanism (such as, e.g., random variation/selective survival of the more intelligent), coupled with a mini-totalitarian state for imputedly stupid survivors.

Perhaps their most strident recommendation is to put more effort into the education of the gifted. After much hand-wringing about efforts to improve the cognitive functioning of the disadvantaged, they conclude that such efforts have largely failed and more resources should now be devoted to a separate, enriched education for high-IQ kids via a voucher system (pp. 435-445, 550). The irony of this recommendation is that the cumulative weight of research now clearly indicates that mixed ability groupings generally benefit the cognitive performance of slower kids without impeding the intellectual progress of brighter ones. In addition, the social development of brighter kids is often enhanced (see, e.g., the review of relevant literature in Curtis, Livingstone, & Smaller, 1992). In spite of Herrnstein and Murray's allusions to generosity and civic mindedness, community, and human dignity, such findings contradict their essential law of competitive individualism. Hence they are ignored.

Lest this commentary be classed as yet another polemic against *The Bell Curve*, let me conclude by noting that I do agree with the authors on some points. We do need to recognize that there are genetic differences in intelligence—but they do not appear to be either as simple or as intractable as Herrnstein and Murray assert. I also agree that there is considerable overlap between blacks and whites, and among social classes, in cognitive ability, and that most people with lower IQs are "good people" (p. 385). But most importantly, I agree that—with the notable exception of my own—the smartest women are not necessarily the best mothers (p. 216).

We do not need more sheer intelligence. We are clearly wasting what we have through underemployment. What the world urgently needs now is more love and compassion across classes, races, sexes, and species. In spite of *The Bell Curve* authors' proclamations of their good intentions, such good cannot come from such an inherently mean-spirited book.

Note

1. The most famous examples are the conscientization projects in which Paulo Freire has been involved. For an impressive US case, see Eberle and Robinson (1980).

References

Cole, E. (1976). *The experience of illiteracy*. Unpublished doctoral dissertation, Union Graduate School, Yellow Springs, OH.

Curtis, B., Livingstone, D.W., & Smaller, H. (1992). *Stacking the deck: The streaming of working-class kids in Ontario schools*. Toronto, ON: Our Schools/Our Selves Educational Foundation.

Duster, T. (1995). The bell curve. *Contemporary Sociology* 24(2). 158-161.

Eberle, A., & Robinson, S. (1980). *The adult illiterate speaks out: Personal perspectives on learning to read and write*. Washington, DC: National Institute of Education.

Esterbrook, G. (1995). Blacktop basketball and *The bell curve*. In R. Jacoby & N. Glauberman (Eds.), *The bell curve: History, documents, opinions* (pp. 30-44). New York: Times Books.

Gould, S.J. (1995). Mismeasure by any measure. In R. Jacoby & N. Glauberman (Eds.), *The bell curve: History, documents, opinions* (pp. 3-13). New York: Times Books.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Lane, C. (1995). Tainted sources. In R. Jacoby & N. Glauberman (Eds.), *The bell curve: History, documents, opinions* (pp. 125-139). New York: Times Books.

Livingstone, D.W. (in press). *The education-jobs gap: Underemployment or economic democracy*. Boulder, CO: Westview Press.

Livingstone, D.W., Hart, D., & Davie, L.E. (1995). *Public attitudes toward education in Ontario 1994: Tenth OISE survey*. Toronto: OISE Press.

Rose, N. (1979). The psychological complex: Mental measurement and social administration. *Ideology and Consciousness 5*, 5-68.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 342-348*

John B. Connors
Canadian Union College

Two Tails are Better than One: The Logic of *The Bell Curve*

With publication of *The Bell Curve* in 1994, Harvard psychologist Herrnstein was once again a controversial figure in the public's eye. Although other recent books had touched on similar issues (Itzkoff, 1994; Seligman, 1992), none had aroused both the ire and praise that Herrnstein evoked. Previous debates with Skinner in 1977 on inherited versus learned behaviors, with Chomsky in 1972 on IQ versus social stratification, and with almost everyone else regarding his matching law in 1961 had made him well known in the scientific community as well. His untimely death in the fall of 1994 left it to his co-author, political scientist Murray, to defend the book on television and radio talk shows (Gunter, 1995).

This article reviews some of the controversies in intelligence testing with particular reference to its impact on Canadian society during the last 100 years. Some of the recommendations made by *The Bell Curve* are compared with earlier proposals made during the eugenics movement at the beginning of the 20th century.

The issue of the role of intelligence and its relation to society is not a new topic for Herrnstein. It was in his now famous article in the *Atlantic Monthly* magazine (Herrnstein, 1971) that he first suggested that as society becomes more egalitarian and artificial social barriers to success are reduced, it inevitably drifts toward a hereditary meritocracy. His support of classical intelligence theory in search of *g* (i.e., a general factor in a hierarchical model of intelligence) is still hotly debated among psychometricans (Azar, 1995). Many other scientists challenge the view that IQ can be a reified construct and argue instead that IQ tests measure only a small sample of behaviors that are more related to social class and educational level than genetics (Gould, 1981; Kamin 1974). Now, in what can arguably be called his magnum opus, Herrnstein revisits the IQ controversy marshalling evidence from a variety of sources (note that the authors themselves did not conduct any of the studies cited) while largely drawing from data in the National Longitudinal Survey of Youth (NLSY) that has followed 12,000 American adolescents and young adults since 1979.

A recent critique of the sources of *The Bell Curve* casts doubt on its scientific credibility (Lane, 1994). Many of the authors cited for psychometric data have been associated with the anthropological journal *The Mankind Quarterly*, which has been a strong supporter of eugenics research and policies. Apparently five of the articles cited in *The Bell Curve* are taken directly from this journal, and 17

John Connors is chair of the Behavioral Sciences Department at Canadian Union College in central Alberta. His research interests include behavior analysis, instructional design, college teaching, and the history of psychology.

of the researchers cited have contributed to it as well. Many of these researchers were also supported by the Pioneer Fund whose benefactors have been supporters of the Nazis. Although this tainted funding of scholars does not by itself invalidate their findings, it does mean that their data should be closely scrutinized. Studies cited that purport to show that Africans score 30 points below whites (which is even lower than American blacks) whereas Asians score 10 points above whites may be due either to test bias or sampling error (Lane, 1994). At this point there is no irrefutable evidence that sets clear biological limits on intelligence.

The nature-nurture debate is an old standard for scientists going back at least as far as the split between Plato and Aristotle in ancient Greece. It has been revived periodically such as in the case of the wild boy of Aveyron with Rousseau versus Itard (e.g., Lane, 1976) at the beginning of the 1800s, in the Terman-Lippmann debates (Block & Dworkin, 1976), and the case study of the Kallikak family by Goddard (Smith, 1985) at the turn of the 20th century. During the 1920s and 1930s the evidence appeared to tilt toward nature and then swung back toward nurture during the 1940s through the 1960s.

An article by Jensen (1969) in the *Harvard Educational Review* rekindled the debate by suggesting that there are real racial differences in intelligence (i.e., a one-standard-deviation difference between blacks and whites) and suggested that little could be done to compensate for these disparate abilities. His evidence was then attacked by many others (Kamin, 1974; Hearnshaw, 1979; Gould, 1981). Jensen continues to counterattack (Jensen, 1980, 1984, 1992) with dense psychometric data to support the nonbiased nature of IQ test scores, whereas opponents contend that genetic bases for intelligence are simplistic and biased and confuse correlation with cause (Lewonton, Rose, & Kamin, 1984; Kamin 1995; Lane, 1994). The nature camp suggests that as much as 80% of intelligence may be directly inherited, whereas the nurture camp is equally forceful is reducing this figure to about 20%. The debates are as often political and ideological as scientific and frequently resort to ad hominem arguments. The authors of *The Bell Curve* offer to split the difference and give an estimate of 60% heritability for intelligence while allowing 40% variation due to environmental factors. Although they admit that dramatic environmental changes could make a difference, they argue that "in reality, what most interventions accomplish is to move children from awful environments to ones that are merely below average" (p. 109). Therefore, the effects of most compensatory programs remain negligible.

Of course the implications of these debates will affect the government social programs in terms of helping both ends of the bell curve: the cognitive elite and the underclass. If the forces of nurture hold sway, then social welfare, affirmative action, and compensatory education are viable and essential components of a democratic society. However, if nature can explain most of the variability in achievement and social class, then government should withdraw help from the underclass and support the cognitive elite who will be the leaders of tomorrow's generation and the only hope for society as a whole. It should be noted that other educators feel that neither better genes nor improved schooling will produce major changes and suggest that true equality can only come about by social and economic changes (Bowles & Gintis, 1976; Jencks, 1972).

As a way of examining the social implications of these academic debates we can look at the eugenics programs that were popular at the end of the 19th century and the beginning of the 20th century. Eugenics tries to improve the human species by using principles of heredity. It includes both the negative eugenics of segregation, sterilization, and euthanasia as well as the positive eugenics of incentives for propagation of the elite (Kelves, 1985). Its goal is essentially selective breeding or what used to be called race betterment.

Although much has been written of the eugenics programs in Britain and the United States (Kelves, 1985), similar programs in Canada have just recently been brought to light (McLaren, 1990). Whereas the British had the problem of their elite migrating to other countries during the 19th century, North America had the problem of accepting multitudes of immigrants. Borrowing a term from population biology, eugenicists warned of dysgenesis, which refers to a downward pressure shift in ability distribution that brings about demographic changes strong enough to have social consequences (Herrnstein & Murray, 1994).

Although Canadian immigration policies usually favored people from other commonwealth countries, many other racial and ethnic groups emigrated to Canada. Europeans were downwardly ranked as either Nordic, Alpine, or Mediterranean. Immigrants were charged with filling up the facilities of asylums, prisons, hospitals, and other charitable institutions while reproducing rapidly. Politically inflammatory statements were made such as "Degenerates among people are worse than bad weeds to a farmer" as well as "Canada had become a dumping ground for the riffraff of the world" (McLaren, 1990, pp. 54, 59). The falling birth rate among Canadians was said to threaten the destiny of the Anglo-Saxon race, with the causes being variously touted as increased urbanism and neurasthenia (a common psychosomatic diagnosis in the 19th century similar to chronic fatigue). Suggestions for change included encouraging people to settle in rural areas, sterilization of the unfit, denying birth control to the fit, and restricting immigration (McLaren, 1990).

Eugenicists at the turn of the century in Canada warned of a direct correlation between the hordes of immigrants and increased insanity, criminality, and unemployment. There was concern not only about large immigrant families that reproduce "like the fish of the sea," but also that the native birth rate was dropping so somehow the newcomers were "sterilizing their hosts" (McLaren, 1990, p. 55). Using the argument that inherited traits were not malleable to a changed environment, eugenicists then attributed all societal ills to the immigrants. The first immigration act in Canada was passed in 1869 and was supposed to screen out lunatics and idiots. Actually, as doctors had no reliable screening methods, they often equated sanity and intelligence with competence in the Anglo-Saxon culture. The Immigration Act of 1910 added to this list in three broad categories:

> The mentally defective included idiots, imbeciles, feeble-minded, epileptics, and insane; the diseased included those afflicted with any loathsome disease or a contagious or infectious disease that might become dangerous to the public health; and the physically defective included the dumb, blind, or otherwise handicapped. (McLaren, 1990, p. 56)

However, it was not possible to implement most of these restrictions due to the following reasons: (a) first-class cabin passengers were not subject to scrutiny; (b) railway and steamship lines were hostile to any operations that reduced their passenger lists; (c) there was not adequate time or staffing to make accurate diagnoses; and (d) the mass of immigrants were understandably not cooperative in providing accurate biographical information (McLaren, 1990).

After World War I Canadian eugenicists had new hope for policing immigration with the establishment of the Department of Health in 1919. Its intended mandate was to improve the efficiency of the population by eradicating feeble-mindedness, infant mortality, tuberculosis, and venereal disease. The waves of immigrants were likened to infectious diseases by calling them a social virus. The Canadian National Committee for Mental Hygiene was also formed in 1918 with the goal of reducing crime, prostitution, and unemployment. However, unlike the United States where mental testing of immigrants at Ellis Island screened out many potential immigrants (the rejection rate in the United States was 1 in 1,500 whereas in Canada it was only 1 in 10,000), physicians had no set standards to work by. In this atmosphere a culprit had to be found that could be more readily identified (McLaren, 1990).

Evidence of the superiority of nature over nurture was established through observations, such as the Army mental testing and the case study of the Dionne quintuplets. Many were convinced that intelligence was passed on mostly by heredity. Henry Goddard, a famous American psychologist, reported that about 50% of all immigrants tested at Ellis Island were feeble-minded (an early term for retardation) and so this classification soon become the source of societal problems (Smith, 1985). An early Canadian eugenicist, MacMurphy (1920), optimistically predicted that 80% of feeble-mindedness could be eliminated by segregation alone, but the ultimate weapon was to be sterilization of the unfit. The call for sterilization actually began in the 1890s due not only to the eugenics movements but also to the perfection of simple operations (vasectomy for men and tubal ligation for women) that were more socially acceptable than castration (Gould, 1985a). There were some cause and effect problems at the time, because some believed that feeble-mindedness caused venereal disease, whereas others believed that venereal disease caused feeble-mindedness. Threats to public health were therefore due to "individual inadequacy" in which feeble-mindedness was the single greatest obstruction to social reform. Case studies of the Kallikak and Jukes families were said to prove that social chaos resulted from reproduction of the unfit over several generations (Smith, 1985). However, the poor were also classified as degenerate and intelligence levels became associated with social class. It was comforting to the middle class that poverty and criminality were due to these individual weaknesses rather than a flaw in the economic structure of the country (McLaren, 1990).

Immigration slowed considerably from about 1914 to 1922 due to the war and a subsequent recession. However, there was much concern during World War I that dysgenesis was still occurring because the fittest men became soldiers and then were killed off while the degenerate were rejected for military service but allowed to breed at home. During the mid 1920s there was an

upsurge in the flow of immigration as the railroad companies sought immigrants to help settle the western prairies. By 1930 it had subsided again due to the depression years and did not pick up again until after World War II. The 1930s can be considered as the heyday of eugenics in Canada due to the scarcity of jobs. During this time sterilization laws were passed in Alberta in 1928 and in British Columbia in 1932 but were defeated in both Manitoba and Ontario, perhaps due to their larger Catholic population who opposed any form of birth control (McLaren, 1990).

Eugenics quickly fell out of favor when the full extent of the Nazi program of racial purification in concentration camps became known. Although strongly opposed by eugenicists, the Family Allowance bill was finally passed by the Canadian federal government in 1945 due to the surplus of jobs available, pressure from Quebec, and the efforts of the welfare-minded in social intervention. Yet the western sterilization laws were not repealed until 1972 (McLaren, 1990). Even today there is a controversy in Canada whether the developmentally disabled should be eligible for major surgery such as organ transplants.

Returning to the logic of *The Bell Curve*, the authors recognize that any attempt on the part of the government to institute negative eugenics policies would make most people apprehensive. Instead they emphasize the government's responsibility to end programs that encourage the underclass. They make their point clearly in one italicized section as follows: "If the United States did as much to encourage high-IQ women to have babies as it now does to encourage low-IQ women, it would rightly be described as engaging in aggressive manipulation of fertility" (p. 548). Besides not subsidizing births the authors advocate that birth control techniques be made increasingly available to women who cannot afford to raise more children. Surprisingly, the whole question of whether underclass women should seek abortions is omitted. Neither are there suggestions on how to encourage the elite to be more fertile. Many social policy suggestions sound like right-wing agenda such as reducing centralized government, more emphasis on the family, tougher punishments on crime, stricter immigration laws, earned income tax credits, and less credentialing of blue-collar jobs.

If none of these government actions is taken seriously, then a scenario is envisioned where the cognitive elite teams up with the already affluent (a total of 5% of the population but which represent 15% of the voting public) and pass laws that keep the underclass in a perpetual welfare state. In this science-fiction society the vices of crime, addiction, and family abuse will be rampant, which will then force the state to take over child care through large orphanages. The homeless will be placed in institutions. New prisons will be built and the police will abuse their new power. Surveillance and control procedures will require national identification cards for all. The underclass will increase in size and be relegated to the inner cities. Racial tensions will be more virulent and may reach a breaking point. Fear of crime may force a totalitarian custodial state, not unlike our current reservations for Native Americans, for the underclass, which would be actively supported by an elite class. In this scenario, both the cognitive elite and the affluent would become increasingly isolated from the underclass while also growing in political and economic power so that they

could bypass most social institutions in favor of private services. Rhetorically, they ask, what will allow these disparate groups to live in harmony? The answer comes in the last chapter entitled "A Place or Everyone." In this new world order, Herrnstein and Murray envision a society where low-IQ people are not pushed to excel but are simply valued for their more menial contributions in an old-fashioned neighborhood where social stratification is expected. Help for the underclass would come through volunteer charities instead of government bureaucracies.

The premise of *The Bell Curve* is that the negative repercussions of the above are due to the ideology of equality. Attempts to equalize humans will lead to inhumane tyrannies that promote diffuse moral outlooks worse than socialism or communism. Public figures are reduced to broaching controversial topics only under the guise of political correctness. The government mentality becomes one of "everything not forbidden is compulsory" (p. 533), so that the moral stance of equality, taken to its extreme, means that it has become unfashionable to talk about any differences between human groups, whether they be black versus white, men versus women, heterosexuals versus homosexuals, young versus old, or high- versus low-IQ groups. Such is the practice of affirmative action that changed from not allowing racial discrimination to requiring equal or equivalent outcomes of racial groups. Many people no longer see the distinction between not interfering and treating everyone the same. And all these views are seen by *The Bell Curve* authors as traditional American values supported by the constitution and the founding fathers. So whether a Malthusian laissez-faire economy and Spencer's survival of the fittest doctrines should prevail at any time in the future may depend on the Zeitgeist that includes the political, economic, and scientific climate of the times.

For Canadians these social implications may sound like an extreme case of the new social order gone amuck. For scientists it must be uncomfortable to view such extrapolations using their own evidence as a wedge. Here the two tails of the bell curve dictate the fate of the rest of the earth. This scenario has, of course, been recently explored in other disciplines such as economics (Reich, 1992). Both positive and negative eugenics have had a recent upsurge in Asian countries such as Singapore, which uses eugenically appropriate computer dating and allows educated women to have more children and gives them priority for enrolling them in the best schools (Gould, 1985b). However, if there is one concept that should have been learned from Darwin it is that the path of evolution—for species, cultures, or governments—is difficult if not impossible to predict. The rules of chaos theory may be more applicable here than classical psychometrics.

So *The Bell Curve* may be neither good science nor good politics. Both tails on the curve appear to be influenced by inaccurate testing and sampling. To suggest that to ignore the eugenics claims is to increase the disparities between these extremes in IQ and thus cause social upheaval may be a misrepresentation of the facts. The middle of the curve will need to remain the dominant force in both numbers and power in a truly democratic society. However, history shows that often scientific achievements are taken out of the hands of scientists when they have political implications (e.g., the Manhattan Project

during World War II to develop a nuclear technology). The track record of the cognitive and economic plutocracies in taking a benevolent dictator role in serving the underclass is not good (e.g., Machiavelli). In this sense *The Bell Curve* seems more of a throwback to the turn-of-the-century eugenics movement rather than a peek at the next century. The skirmish that Canadians had with eugenicists and their social policy recommendations in the 1920s and 1930s should make them wary of using any one criterion for social mobility. The two new tails of the bell curve may now be the scientists against the politicians.

References

Azar, B. (1995). Searching for intelligence beyond "g." *APA Monitor, 1*, 25.

Block, N.J., & Dworkin, G. (Eds.). (1976). *The IQ controversy*. New York: Pantheon Books.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America*. New York: Basic Books.

Gould, S.J. (1981). *The mismeasure of man*. New York: Norton.

Gould, S.J. (1985a). Carrie Buck's daughter. In S.J. Gould (Ed.), *The flamingo's smile* (pp. 306-318). New York: Norton.

Gould, S.J. (1985b). Singapore's patrimony (and matrimony). In S.J. Gould (Ed.), *The flamingo's smile* (pp. 319-332). New York: Norton.

Gunter, L. (1995, February). Tax dollars can't buy you brains. *Alberta Report*, pp. 42-43.

Hearnshaw, L.S. (1979). *Cyril Burt, psychologist*. New York: Vintage Books.

Herrnstein, R.J. (1971, September). IQ. *Atlantic Monthly*, 43-64.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Itzkoff, S.W. (1994). *The decline of intelligence in America: A strategy for national renewal*. New York: Praeger.

Jencks. C. (1972). *Inequality: A reassessment of the effect of family and schooling in America*. New York: Basic Books.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39*, 1-123.

Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.

Jensen, A.R. (1984). Test bias: Concepts and criticisms. In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on "Bias in Mental Testing* (pp. 507-586). New York: Plenum Press.

Jensen, A.R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research, 27*, 225-233.

Kamin, L.J. (1974). *The science and politics of IQ*. New York: Wiley.

Kamin, L.J. (1995). Behind the curve [Review of *The bell curve: Intelligence and class structure in American life*]. *Scientific American, 272*(2), pp. 99-103.

Kelves, D.J. (1985). *In the name of eugenics*. Berkeley, CA: University of California Press.

Lane, C. (1994). The tainted sources of *The Bell Curve* [Review of *The bell curve: Intelligence and class structure in American life*]. *New York Review of Books, 41*(20), 14-19.

Lane, H. (1976). *The wild boy of Aveyron*. Cambridge, MA: Harvard University Press.

Lewontin, R.C., Rose, S., & Kamin, L.J. (1984). *Not in our genes*. New York: Pantheon Books.

MacMurphy, H. (1920). *The almosts: A study of the feeble-minded*. Boston, MA: Houghton-Mifflin.

McLaren, A. (1990). *Our own master race: Eugenics in Canada, 1885-1945*. Toronto, ON: McClelland & Stewart.

Reich, R.B. (1992). *The work of nations*. New York: Vintage Books.

Seligman, D. (1992). *A question of intelligence: The IQ debate in America*. New York: Carol Publishing Group.

Smith, J.D. (1985). *Minds made feeble: The myth and legacy of the Kallikaks*. Rockville, MD: Aspen Systems.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 349-354*

Catherine D. Krull

and

W. David Pierce
University of Alberta

IQ Testing in America: A Victim of its Own Success

A major thesis of *The Bell Curve* (Herrnstein & Murray, 1994) is that intellectual capacity is concentrated into social classes. According to Herrnstein and Murray, this was not always the case. At an earlier point in United States history, few specialized occupations required intellectual competence. Class structure was based on wealth, with intelligence being evenly distributed among the classes. As the US changed toward a "high tech," knowledge-based economy, it became important to identify individuals with superior cognitive ability. The authors of *The Bell Curve* make it clear that mass education and IQ testing were the solutions to US economic requirements.

In this regard we note that mass education ensured that most people went to school regardless of background. That is, mass education fitted well with the US principle of equal opportunity. Ironically, once the masses were going to school, the focus changed from educating all students to identifying and selecting the bright ones through intelligence testing (so much for equal opportunity).

A point not mentioned in *The Bell Curve* is that IQ tests were originally used to monitor and assess the effectiveness of the educational system (for a review of educational testing, including IQ, see Haney, 1984). That is, early educators hoped to develop curriculum and programs that suited the learning needs of students regardless of social background. What has emerged, however, is an educational system directed at performance on IQ tests (e.g., scholastic aptitude tests). The development of effective educational programs has become secondary. In the present system the bright students are able to flourish while other students fall by the wayside (so much for mass education).

From our perspective, IQ testing has contributed to the low involvement of US parents in the formal education of their children.[1] A score on an IQ test is often taken as a measure of a child's innate cognitive ability. This assumption implies that parents can do little to enhance their children's academic performance. Parents in the US tend to set standards of performance to meet their child's IQ level rather than to encourage effort to overcome learning difficul-

Catherine Krull is a doctoral candidate in the Department of Sociology. Her interests include social demography, education, and gender issues.
David Pierce is a professor in the Department of Sociology, and an adjunct professor in the Department of Neuroscience. His interests include behavioral choice and preference, applications of behavior principles to culture and education, and a biobehavioral theory of activity anorexia.

ties. In other words, IQ scores result in a self-fulfilling prophecy: parents (and teachers) believe their children cannot learn beyond a certain point and children meet these expectations.

As Herrnstein and Murray document, and we concur, the overall result of US IQ testing has been the rise of a class structure based on intelligence, with dull people on the bottom and the cognitive elite on the top. Based on the IQ-selection process, bright people have become concentrated in prestigious, knowledge-based occupations that yield high income and wealth. Conversely, the intellectually dull have been relegated to low-prestige jobs outside the information economy. Today many of these people are often unemployed.

Herrnstein and Murray also argue that the dull underclass is contributing to most of the US's woes. At least half their book concerns the correlation between IQ and a wide range of apparently intractable social problems. They describe the correlation between a high incidence of the social problem and the low intelligence of those who perpetuate it. This correlation is shown to hold for school dropouts, unemployment, crime, out-of-wedlock births, poor parenting, and many other social ills. We show that the relationship between IQ and social problems has been manufactured and is amenable to change through reform of US education.

The authors of *The Bell Curve* argue that many of us (mostly liberals) refuse to acknowledge the relationship between intelligence and social problems. However, according to the authors, the nation can only address its problems if it first faces up to the undeniable news. Americans must accept immutable differences in intelligence by groups of people, as well as similar differences among individuals in any group. Because differences in intelligence are unalterable, there is no point in special Head Start programs for underprivileged youth. Herrnstein and Murray conclude that the IQ gains of such programs are seldom realized and never maintained. Also, programs such as affirmative action are unjust and contribute to inefficiencies on the job and lower productivity in industry. Although the authors competently outline the problems that have resulted from intelligence testing, they offer no solutions. They merely conclude that solutions might be found if Americans had the courage to talk about these issues in public. Our position is that Americans would do better to discuss and set new educational objectives than to dwell on the IQ controversy.

Manufacturing Social Problems: IQ testing and Selection

As mentioned above, in chapter after chapter Herrnstein and Murray document the correlations between IQ and social ills. Interestingly, according to them, these correlations did not exist in the US before IQ testing was used as a means of concentrating individuals into social classes (see, e.g., the 1960s rise of crime, p. 236). In other words, a hidden message in *The Bell Curve* is that the IQ-selection process has resulted in engineered correlations between IQ scores and diverse social problems. For example, the selection of students based on IQ scores by elite US universities ensures that only a small percentage of people can rise to the top. Individuals who have average scores must attend less prestigious universities or go to city colleges. Those who score low on IQ tests either drop out of high school or merely complete their high school education. Consequently, they are assigned to the unskilled sector of the labor force. Thus the correlations among IQ and social problems such as unemployment or

poverty (and other associated social ills) are manufactured by the IQ-selection process.[2]

We agree with Herrnstein and Murray that US society indeed has been transformed to its detriment by concentrating people with similar IQ levels into distinct social classes. The authors, however, assume a status quo stance on the significance of IQ testing for US education. At no time do Herrnstein and Murray consider the possibilities of eliminating or reducing the use of IQ tests to determine achievement. Analysis of the system, however, immediately suggests that the source of US woes lies with an educational institution that accepts and propagates intelligence testing. From our perspective, then, the US's problems lie in a culture that emphasizes innate abilities and deemphasizes achievement through learning.

In addition to emphasizing the manufactured correlations between IQ and social problems, Herrnstein and Murray repeatedly show that socioeconomic status (SES) has little effect. Liberal thinkers may find it surprising that the correlations between IQ and social problems are not substantially reduced by taking SES of parents into account. That is, the kind of home background (well off or impoverished) does not predict well the incidence of various social problems. The authors of *The Bell Curve* would have us believe that SES has weak effects because IQ and social problems are inherently linked. Herrnstein and Murray do not consider the possibility that all correlations among IQ, SES, and social problems are engineered. That is, it is quite possible that the US system of IQ testing and selection of the smart (rejection of the dull) has reduced the impact of social background on indicators of Americans' way of life. In fact, one of the goals of educators and IQ testers was to allow gifted children to rise to the top independent of their social position. As with the correlations between IQ and social problems, the weak association between social background and social ills has been contrived by the education and testing systems.

The important point of our analysis is that there is no immutable relationship between IQ and the US's social problems. If correlations have been manufactured by a social system, it is possible to use social planning to ameliorate these relationships. One obvious step is to eliminate or reduce the impact of IQ testing as a primary basis for selection into higher education. This involves taking into account the diverse talents, skills, and motivations of individuals in the educational assessment and selection process. Another step is to modify US values and practices toward a culture of learning, where most people are expected to succeed in academic subjects, rather than only the gifted few.

America's Shame: Testing versus Learning

Equal opportunity for achievement is ingrained as a fundamental part of the US's values. The concept of equal opportunity means that individuals are presumed to have similar chances of achievement and success. Individual decisions based on skills, motivation, and performance allow one person to advance while others fall behind. For example, the entrepreneur who identifies business opportunities, arranges for investment of capital, and works hard to achieve her or his goal, is expected to succeed. Other persons who could also

pursue business ventures but do not seize the opportunity, take the risk, or put out the effort are expected to have less success.

Equal opportunity presupposes a meritocracy where each person competes and advances on the basis of performance. Unfortunately, IQ testing and selection has limited the competition for success to a single indicator of worth; merit is not assessed by actual performance, but assigned on the assumption that IQ scores measure the inherent general intelligence of people (the so-called *g* factor). All other facets of personality, motivation, and skills are underplayed. With Herrnstein and Murray's view that IQ scores are immutable and transmitted at birth from generation to generation, the majority of Americans continue to be excluded from the competition for prestige positions.

In order for the US to become a true meritocracy, a majority of people must be equipped with the skills, knowledge, and motivation to succeed in life. Such a change requires a shift in education away from testing for individual differences, toward a system of learning and performance. At the most basic level there would be a change from measuring reasoning capacity to an acceptance that reasoning is a skill that can be acquired by the majority of students.

The idea that reasoning is an acquired skill, rather than an immutable attribute, is illustrated by the cross-national comparisons of mathematics achievement by Stevenson and his colleagues.[3] There is little doubt that learning mathematics equips people with abstract analytical skills and at the same time allows them to solve many practical problems. Also, the information, high-tech world economy demands the skills of mathematics and analytical reasoning. Thus an American society based on equal opportunity and merit would do well to endow its citizens with mathematical expertise. Surprisingly, mathematical achievement in the US is substantially below the performance levels of Asian countries.

An important consideration, given Herrnstein and Murray's emphasis on IQ and achievement, is the finding that children's cognitive ability does not differ among China, Japan, and the US.[4] That is, the results indicate that the high mathematical achievement of the Chinese and Japanese students is not due to their greater intellectual capacities. The differences in achievement are, however, related to cultural practices in the areas of education and learning. For example, in terms of classroom management and organization in the US, Stigler et al. (1987) report:

> American children fail to receive sufficient instruction. They spend less time each year in school, less time each day in classes, less time in the school day in mathematics classes, and less time in each class receiving instruction. The classes were organized so that American children were frequently left to work alone at their seats on material in mathematics that they apparently did not understand well, they engaged in many irrelevant activities, and they spent large amounts of time in transition from one activity to another. (p. 1285)

These classroom difficulties must be understood in a broader cultural context. In contrast with Americans, Asian cultures emphasize effort as the basis of individual achievement and minimize the importance of intellectual capacity. This concern with effort and achievement means that Chinese and Japanese children are expected to work hard to learn academic subjects and that teachers and parents are expected to set high standards for performance. Based on these

beliefs, Asian children spend substantially more time in school on mathematics and have greater homework assignments than US students. Students and teachers "are more intensely involved in mathematics than are their American counterparts. Chinese and Japanese teachers appear to be better prepared for teaching mathematics and endow their classes with liveliness and variety that typically are missing in American elementary school classes in mathematics" (Stigler et al., 1987, p. 1284). In addition, Asian children are expected to strive to learn the subject matter and to overcome academic problems by spending more time on difficult material. In Asian cultures parents are expected to arrange a study space for their children, supply additional educational materials as needed, and spend time with their children on academic problems.

The point is that people, regardless of IQ differences, show high achievement in mathematics when a culture emphasizes effort and learning rather than immutable cognitive ability. An obvious solution to Herrnstein and Murray's intelligence and social class thesis (and the associated social problems) is for the US to adopt values and practices that mirror those countries with high mathematical achievement. In order to accomplish this, there would have to be a standardized curriculum in core subjects throughout the US. This would eliminate local variation in mathematical instruction (or other subjects). In this system teachers would have high status and be proficient in their subject matter. Children, on the other hand, would be expected to learn the academic material and to work hard on difficulties. Parents would assume partial responsibility for their children's learning, aid in homework and self-study, set high but attainable standards for achievement, and evaluate their children's progress toward short- and long-term academic objectives. Overall, the US would change from a culture of testing toward a culture of learning—with most people, rather than the few with high IQ, prepared by their education to compete for the opportunities of life.

Americans have notoriously been known for their belief in individualism, yet they have denied themselves a system where an individual can rise to the top through effort and hard work. It would appear that the American Dream is more of a myth today than in previous generations due to the emphasis placed on IQ testing rather than on personal achievement through effort and learning.

A Canadian Footnote

Our analysis also has important implications for the Canadian educational system. Although Canada has not implemented IQ testing to the same degree as the US, Canadian students have done just as poorly as their US counterparts on mathematical tests. In the Canadian system education is tailored to ensure that most students pass, with less attention paid to effort and achieved competence in the subject matter. In order to be certain that most people make it through the system, the content of the curriculum is continually downgraded to meet the levels of striving and achievement of Canadian students. Importantly, low levels of mathematical achievement in Canada could be ameliorated by the educational system we have outlined. Canadian parents and teachers, like their US counterparts, tend to set low standards for achievement and are often satisfied with average effort in core subjects. Moreover, as Stevenson's studies demonstrate, students in Canadian schools place more responsibility for learning on the teacher than on themselves.[5] That is, the

students attribute their mathematical performance to external sources (i.e., teachers) rather than personal causes such as effort and motivation. This attributional style means that Canadian students do not strive for achievement in mathematics and science and blame their failures on others. The overall pattern of attribution and expectations of Canadian students has severe implications for their future prospects. Compared with work-oriented students in other countries, Canadian students do not have (and do not seek) the high-level skills in mathematics and science that are valued by the global market. Without a change in our practices related to education, students from both the US and Canada will continue to lose ground in the world competition for knowledge-based employment. The survival of North American culture is at stake. An educational system based on effort and learning is the answer.

Notes

1. More information on parent involvement in the formal education of their children appears in Stevenson, Lee, Chen, Stigler, Hsu, and Kitamura (1990).
2. We are, of course, arguing that the relationships among IQ and social problems are spurious rather than causal. In our analysis, the covariation of IQ and any social problem results from the IQ-selection process (the causal factor).
3. For further information on mathematical achievement, see Crystal, Chen, Fuligni, and Stevenson (1994); Crystal and Stevenson (1991); Stevenson, Chen, and Lee (1993); Stevenson, Lee, Chen, and Lumis (1990); Stevenson, Lee, and Stigler (1986); Stevenson et al. (1990); Stevenson, Stigler, Lee, and Lucker (1985); Stigler, Lee, & Stevenson (1987).
4. Stevenson et al. (1985) report that the high achievement of Chinese and Japanese children cannot be attributed to higher intellectual abilities, but must be related to their experiences at home and at school.
5. The mathematical achievement results for Canada were presented by Harrold Stevenson to the Faculty of Education at the University of Alberta in 1994. Our discussion of Canadian education and mathematical acheivement is based on summaries of these results.

References

Crystal, D.S., & Stevenson, H.W. (1991). Mothers' perceptions of children's problems with mathematics: A cross-national comparison. *Journal of Educational Psychology, 83*, 372-376.

Crystal, D.S., Chen, C., Fuligni, A., & Stevenson, H.W. (1994). Psychological maladjustment and academic achievement: A cross-cultural study of Japanese, Chinese, and American high school students. *Child Development, 65*, 738-753.

Haney, W. (1984). Testing reasoning and reasoning about testing. *Review of Educational Research, 54*, 597-654.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Stevenson, H.W., Chen, C., & Lee, S. (1993). Mathematics achievement of Chinese, Japanese, and American children: Ten years later. *Science, 259*, 53-58.

Stevenson, H.W., Lee, S., Chen, C., & Lumis, M. (1990). Mathematics achievement of children in China and the United States. *Child Development, 61*, 1053-1066.

Stevenson, H.W., Lee, S, & Stigler, J.W. (1986). Mathematics achievement of Chinese, Japanese, and American children. *Science, 231*, 691-699.

Stevenson, H.W., Stigler, J.W., Lee, S., Lucker, G.W. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56*, 718-734.

Stevenson, H.W., Lee, S., Chen, C., Stigler, J.W., Hsu, C. & Kitamura, S. (1990). Contests of achievement: A study of American, Chinese, and Japanese children. *Monographs of the Society for Research in Child Development, 55*.

Stigler, J.W., Lee, S., & Stevenson, H.W. (1987). Mathematics classrooms in Japan, Taiwan, and the United States. *Child Development, 58*, 1272-1285.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 355-359*

Michael Schleifer
Université du Québec à Montréal

Should We Change our Views About Early Childhood Education?

The Bell Curve (Herrnstein & Murray, 1994) has facets too numerous for one review. Clearly each commentary must restrict itself to one subtheme of the book. My own choice is the consequences (implied or stated) for early childhood education. If I had chosen a second theme, it would have been intelligence and racism. Before discussing my main theme, here are a few comments on this other issue.

I do not believe that the defense against accusations of racism is nearly strong enough; the burden of proof remains today, as it did 20 years ago, on those who pursue research into differences between blacks and whites. The motivation for having undertaken this kind of research project—if other than based on a racist agenda—was not sufficiently forthcoming when Jensen presented his research in 1969. Nor does the present book come close to allaying these same suspicions. The disclaimers in the text are in the right direction; they are undermined, however, by other comments. One example concerns the research of Burt, a major player in IQ history. The allegations of falsification of research data are treated somewhat cavalierly, in the form of a small text box (p. 12) that simply repeats that some researchers side with Burt against his detractors. Given the explosiveness of the subject matter, I expected much more detail and argument. We are left at the most with an impression of "not proven," rather than "not guilty." Equally feeble is the reminder that Burt's research conclusions concerning correlations between twins reared separately have been found by others. A second example of the cursory treatment afforded these issues appears in the Introduction (p. 5) where the accusation of racism in regard to early immigrants is treated. Once again, the box feebly attempts to downplay the issue. At best, there is an attempt once again to make the case of "not proven." The tone in each of these examples (there are numerous others throughout the text) is nowhere serious enough. One gets the impression that lip service is being paid to the racism issue, but no more.

My focus in the present commentary, as mentioned above, is education: more particularly preschool education and the child from birth to 5 years old. Consensus in Quebec is that offering education at this age is a good thing, and not enough of it is being done. Compensatory education, or enrichment programs, are seen as particularly beneficial to the development of children suffering from one or other deficits. They are also seen as offering hope for raising the level of intellectual functioning of the vast majority of children.

Michael Schleifer, a graduate of Oxford University and McGill, teaches in the Education Department. His main research interest is moral and logical reasoning.

Should any of these views be altered in view of the findings and arguments of *The Bell Curve*? The authors themselves are ambiguous concerning the educational consequences. At times they wish to adopt the view that there are no such consequences. Their thesis leaves everything as it is: "If tomorrow you knew beyond a shadow of a doubt that all the cognitive differences between races were 100 percent genetic in origin, nothing of any significance would change" (p. 314).

This view, reminiscent of the apologetics of Jensen in his 1969 article, seeks shelter in a form of the *naturalistic fallacy*. Thus the findings about intelligence are facts, and the educational conclusions are *oughts*. According to the philosopher David Hume, no *is* can imply an *ought*. Hence everything remains the same. To deduce an *ought* from an *is* is to commit a form of the naturalistic fallacy. This philosophical ploy is itself controversial, however. This is both because Hume and the naturalistic fallacy are debatable and debated (Flew, 1973; Hudson, 1969; Schleifer, 1973). In the context of intelligence and education, moreover, the lines between facts and values, between *is* and *ought* are even fuzzier than elsewhere. The authors, in their history of IQ, chose not to mention that Binet's test was for both intelligence and educability. These terms were used interchangeably in the context of having to decide which children would go to school and which not (there were not sufficient available places in France at the time).

In any case the authors themselves sometimes retreat from this nothing-follows line. They offer the contradictory and no less controversial version that does trace educational consequences. Here the thesis does attack the consensus mentioned above that compensatory education is worthwhile. First the authors remind us of differences between races on specific measures of intelligence:

> The universality of the contrast in nonverbal and verbal skills between East Asians and European whites suggests, without quite proving, genetic roots. Another line of evidence pointing toward a genetic factor in cognitive ethnic differences is that blacks and whites differ most on the tests that are the best measures of g, or general intelligence. (p. 270)

Herrnstein and Murray mention spatial-perceptual ability (p. 302) citing Jensen (1969). Jensen had himself lumped these various abilities into abstract reasoning abilities. The first part of the argument, then, is to remind us of the genetic component linked to race, with the implication not to waste time or funds on inappropriate education. The second part of the argument is to be found in Chapter 17, entitled "Raising Cognitive Ability." Here we have the reaffirmation that Head Start and similar programs were a failure, and all will necessarily fail. Trying to bolster cognitive capacity wastes funds because it can only have a small effect, and for a short term (p. 389). Ignored here are those studies (Anderson & Messick 1974; Miller & Dyer, 1970, 1975) that have shown long-term improvement on intellectual persistence and effort when early intervention concerns itself with attitudes, motivation, and the total child.

Most of the studies and findings cited by the authors to illustrate the failure of compensatory and enrichment programs depend on their notion of cognitive ability that, as noted above, is reduced to intelligence, and then to IQ. Other studies (Boutin & Terrisse, 1990; Casto & Mastropiere, 1986; Dunst, Snyder, & Mankein, 1987) find exactly the opposite, namely, that preschool

programs have a long-term effect on compensating for early deficits. The crucial ingredient in these programs seems to be the implication of parents (Isserlis, McCue, Weinstein, & Sauvé, 1994; Terrisse & Boutin, 1994). Along these lines, several ongoing pilot projects are currently being conducted in lower-income areas in Quebec, British Columbia, and various sites in the United States offering early childhood programs, with the parents very much involved. Ironically, the authors of *The Bell Curve* themselves offer pertinent data showing the importance of the level of functioning of parents to the development of their children (pp. 220-232). Perhaps because of their own agenda they fail to see that these findings support, rather than oppose, the relevance of early education, particularly of the kind outlined above.

Another ingredient ignored by the authors in their evaluation of Head Start involves nutrition. Although on the one hand they themselves offer (pp. 389-391) data to support the importance of good nutrition in bolstering cognitive ability (even in their limited sense of IQ), they fail once again to draw the consequences for the beneficial effects of early childhood programs. As early as 1973, assessment studies of some Head Start programs had noted the improvements in children's health, nutrition, and social outlook (Richardson & Spears, 1972). It has been well documented that the participation of blacks in political life increased dramatically from 1963 on. These changes are widely acknowledged as due to the operation of Head Start programs. The authors are happy to mention ill health and poverty as important factors, but fail to draw the link between these two and early education. This link does manifestly exist.

Those who work with parents and young children believe that cognitive ability is enhanced. The studies, furthermore, that demonstrate the positive gains in these programs make use of a notion of cognitive ability different from that of Herrnstein and Murray. In the first place, aspects of cognitive, social, and emotional development are intertwined and linked. Children of 3 and 4, for example, who reflect on the notion of a rule are more likely at the age of 5 and 6 to understand a rule (cognitive development), internalize certain rules previously linked to an authority (social development), and accept certain rules with positive affect (emotional development). This research perspective is inspired by the work of Piaget.

The authors themselves too quickly dismiss a paradigm about cognition associated with Piaget. His view, relegated to the "revisionists," is seen as having a "different focus" (p. 20). They prefer, of course, the view about intelligence that permits the traditional IQ test and reduces cognitive ability and intelligence to IQ. What is important about education, however, is that Piaget's developmental view retains a strong concept of cognition without reducing this to changes in IQ. What the authors have dismissed as a change of focus is much more. The Piagetian view of cognitive development and cognitive ability is diametrically opposed to their notion of IQ. For Piaget intelligence is a continuous process of individual growth; it is not a set of discrete capacities. For Piaget the tests of Binet and Burt were necessarily limited, because he was interested in justifications and reasoning in order to better understand mental structure. Once again, this is much more than a difference in focus; it is an opposing view. Finally, the biological component of the Piagetian perspective must be mentioned in contrast to the psychometric ap-

proach that is based on the model of physics. Thus for Piaget there is always a structure evolving by interaction with the environment into a more complex structure. The nature-nurture debate is seen differently from this point of view. The other facet of the biological perspective is that the organism is essentially organized, not just a set of parts making up a whole. The affinity of Piaget for the integral personality, and education of the whole person, as seen in John Dewey's work, can be understood in the light of these biological underpinnings. Explicitly or implicitly, many of these Piagetian facets have influenced educators. For example, the Curriculum for Preschool Education of the Quebec Ministry of Education offers as its overall objective "To allow the preschool child to pursue his [or her] own path, to encourage his abilities, to develop relationships with others, and to interact with his environment" (p. 10).

This objective of preschool education is coherent with the view outlined above in regard to development adopted in research. It should be noted that early childhood education in this sense looks at the whole person and views the child as a person who acts and interacts with his entire being. It should be noted also, that this view of early childhood education applies to the functioning of the great majority of normal children, as well as those in need of specific enrichment due to impairment or deficit. Contrary to the authors' suggestions that only the elite ("those with the greatest potential") can and should be offered improved education (pp. 445-550), the majority of us working in education continue to believe in the appropriateness of aiming at educating the majority. I am working at the moment with a group of teachers who are introducing logic and philosophy at early ages in schools, including kindergartens and nurseries. One of the most visible results involves the exchanges, conversations, and arguments between children of all language backgrounds, colors, ethnic groups, and social milieux. What Herrnstein and Murray suggest as impractical and a waste, I see as the most precious of investments one can offer. At the research level, longitudinal studies are underway to evaluate the impact of the program. Needless to say, the criteria for assessing cognitive ability and cognitive development will not be in terms of IQ.

References

Anderson, S.B., & Messick, S. (1974). Social competency in young children. *Developmental psychology, 10*, 282-293.

Boutin, G., & Terrisse, B. (1990). Programs d'intervention précoce. In S. Jonescu (Ed.), *L'intervention en déficience mentale, Vol. 2, Manuel de méthodes et de techniques* (pp. 15-41). Brussels: P. Mardaga.

Casto, G., & Mastropieri, M.A. (1986). The efficacy of early intervention programs: A meta-analysis. *Exceptional children, 52*, 417-424.

Dunst, C.J., Snyder, S.W., & Mankein, N. (1987). Efficacy of early intervention. In M. Wang, M.R. Clinton, & H.J. Walberg (Eds.), *Handbook of special education: Vol 3, Research and practice: Low evidence conditions* (pp. 259-293). New York: Pergamon.

Flew, A. (1973). The Jensen uproar. *Philosophy, 48*(183), 63-69.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Hudson, W.D. (Ed.). (1969). *The is/ought question*. London: Macmillan.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39*(1).

Isserlis, J., McCue, L., Weinstein, L. & Sauvé, V. (1994). *Community Literacy: An intergenerational perspective*. Sponsored by the National Literacy Secretariat.

Miller, L.B., & Dyer, J.L. (1970). *Experimental variation of Head Start curricula, a comparison of current approaches*. Louisville, KY: Child development Laboratory, University of Louisville.

Miller, L.B., & Dyer, J.L. (1975). Four preschool programs: Their dimensions and effects. *Monographs of the society for research in child development, 40* (Series No. 162, Nos. 5-6).
Richardson, R., & Spears, D. (1972). *Race, culture and intelligence*. Harmondsworth, UK: Penguin, 1972.
Schleifer, M. (1973). The Flew-Jensen uproar. *Philosophy, 48*, 386-390.
Terrisse, B., & Boutin, G. (1994). *La famille et l'éducation de l'enfant de la naissance à six ans.* Montréal, PQ: Les Éditions Logiques.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 360-366*

David G. Wangler
University of Alberta

Is *The Bell Curve* a Ringer?

In reading, thinking, or writing about *The Bell Curve* (Herrnstein & Murray, 1994) it is important, perhaps critical, to note that the book is not primarily about race or what the authors call ethnic differences (p. 271). The first sentence of the preface to the book states "this book is about differences in intellectual capacity among people and groups and what those differences mean for America's future." The book itself excluding the appendixes is 527 pages long. Roughly 70 pages (269-340) are devoted exclusively to the relationship between race and IQ. About 14% of the actual text considers the relationship between IQ levels and racial/ethnic background. Despite the fact that 86% of the book considers IQ primarily in reference to class, it is the three chapters out of a total of 22 that deal with the IQ-race relationship that have received the most attention from critics on both the right and the left.

There are large problems with the racial ethnic comparisons that Herrnstein and Murray make in the chapter that compares ethnic differences in cognitive ability. Herrnstein and Murray compare blacks and whites, but they note that they categorize people on the basis of "whatever they prefer to be called" (p. 271). These comparisons suggest that there are distinct racial groups and that these groups differ on many variables including IQ. It is obvious that a black person looks different than a white person, but is there a pure black or pure white group? Are we comparing somewhat, largely, or completely black people with white individuals who are qualified in a similar manner. If yes, how do we label or categorize someone who is half white and half black? A recent book, *The History and Geography of Human Genes* by geneticists Cavalli-Sforza, Menozzi, and Piazza (1993), suggests that after we get by surface traits such as skin color or hair texture there are very few significant differences among human groups. They note, as have many others including Murray and Herrnstein, that there is much greater variation within groups than there is between groups in terms of differences. Their conclusion is that at the level of genes, the very idea of race is lacking in significance. Now if at the genetic level we are remarkably similar, how can Murray and Herrnstein argue for genetically based differences in IQ? If Cavalli-Sforza et al. are correct, and there is much support for their position in modern anthropological and demographic studies, then the idea that there are substantial genetic differences between races is either suspect or false. How can one compare IQ scores according to a category that does not exist?

Herrnstein and Murray are aware of this problem and even cite Gould's observation that

David Wangler has been a professor in the Department of Educational Policy Studies for 26 years and teaches in the area of history of childhood and the senior elective course Issues in Education.

> We now know that our usual metaphor of superficiality—skin deep—is literally accurate. Say it five times before breakfast tomorrow; more important, understand it as the center of a network of implication: "human equality [i.e., equality among the races is a contingent fact of history]." (p. 296)

The authors of *The Bell Curve* try to get around this problem by writing "but some ethnic (racial) groups nonetheless differ genetically for sure, otherwise they would not have differing skin color or hair textures or muscle mass. They also differ intellectually on the average" (p. 297). But we are still faced with the problem of which ethnic/racial category to put someone in who looks black but whose genes are 50% white. To complicate matters further, Europeans, who are the primary representatives of the white group in *The Bell Curve*, are believed to be a hybrid group made up of 65% Asian and 35% African genes. Thus the designated white group in this book is anything but white.

The word *racist* has been used extensively to describe both the book, the three chapters dealing with racial differences, and the authors themselves. It is probably not beneficial or helpful to label people we disagree with as being racists or rednecks or bleeding heart liberals or limousine socialists. It is indeed possible that the people who present positions that are contrary to or in direct opposition to our own may not only have something to say but may possess aspects of the truth that we have overlooked or ignored. In listening to and considering their positions we may well strengthen or understand our own position better. This was eloquently said by John Stuart Mill (Cohen, 1961) in his "Essay on Liberty" wherein he wrote that

> The peculiar evil of silencing the expression of an opinion is, that it is robbing the human race; posterity as well as the existing generation; those who dissent from the opinion, still more than those who hold it. If the opinion is right, they are deprived of the opportunity of exchanging error for truth: if wrong, they lose, what is almost as great a benefit, the clearer perception and livelier impression of truth, produced by its collision with error. (p. 205)

Mill goes on to suggest that stifling an opinion is evil even if it is false.

> First: the opinion which it is attempted to suppress by authority may possibly be true. Those who desire to suppress it, of course deny its truth; but they are not infallible. They have no authority to decide the question for all mankind, and exclude every other person from the means of judging. To refuse a hearing to an opinion, because they are sure that it is false, is to assume that their certainty is the same thing as absolute certainty. All silencing of discussion is an assumption of infallibility. Its condemnation may be allowed to rest on this common argument, not the worse for being common. (p. 205)

The Bell Curve presents many ideas that are contrary to majority opinion and especially to mainstream social science perspectives. It is a book that should be carefully read, debated, and discussed. Its evidence should be extensively analyzed and weighed and, if and where possible, complemented or refuted. The authors should not be regarded as Neo-Nazis (Rosen & Lane, 1994, p. 14) as one critic labeled them, nor should their work be seen as sleazy and an intellectual mess, as Ryan (1994, p. 11) suggests. Any attempt to deal with controversial social or individual problems has to be done in a candid, forthright manner that is devoid of namecalling or mudslinging. This has been the exception rather than the norm in the numerous reactions to this book.

What is desperately needed if we are to overcome our own ignorance and achieve some modicum of truth (notice the small t) is contrary evidence of a rational, empirical, or logical nature. Perhaps any discussion of class or racial differences in reference to IQ or any other variable like race or class should be guided by Pascal's observation that "we know too little to be dogmatic and too much to be skeptical."

With this in mind it is necessary to consider the significance or value of a book like *The Bell Curve*. The book contains a wealth of information on a number of different topics. It is difficult not to be impressed by the range of ideas and information presented and by the large number of graphs, charts, and illustrations used to clarify and condense the material presented. If you're into statistics, charts, and graphs as a way of concisely presenting data and results, then this book is for you.

On the general and primary theme of the relation of intelligence to social status and social pathology, *The Bell Curve* clearly demonstrates that being in the bottom one or two deciles of the IQ range of the population is often associated with high levels of social pathology. To quote the authors, "We have tried to point out what a small segment of the population accounts for such a large proportion of those problems" (p. 549). The problems referred to in the quote are things like illegitimacy: "The knowledge that 95 percent of poor teenage women who have babies are also below average in intelligence should prompt skepticism about strategies that rely on abstract and far sighted calculations of self interest" (p. 387). Poverty is also noted: "The high rates of poverty that affect certain segments of the white population are determined more by intelligence than by socioeconomic background" (p. 141).

The authors acknowledge that the level of education one achieves is related to social class, but then note that "if cognitive ability is high, socioeconomic disadvantage is no longer a significant barrier to getting a college degree" (p. 154).

In terms of race, intelligence, and level of education, *The Bell Curve* points out that for those possessing the 103 average IQ of all high school graduates, the chance of completing high school went up if one was black or Latino:

> Consider, for example, graduation from high school. As of 1990, 84 percent of whites in the NLSY had gotten a high school diploma, compared to only 73 percent of blacks and 65 percent of Latinos, echoing national statistics. But these percentages are based on everybody, at all levels of intelligence. What were the odds that a black or Latino with an IQ of 103—the average IQ of all high school graduates—completed high school? The answer is that a youngster from either minority group had a higher probability of graduating from high school than a white, if all of them had IQs of 103: The odds were 93 percent and 91 percent for blacks and Latinos respectively, compared to 89 percent for whites. (p. 319)

In opposition to many sociologists, Herrnstein and Murray see almost all the variables noted above as relating to or emanating from intelligence. Their position is that low intelligence is an important factor in where the individual finds himself or herself in the social class structure of the United States. To use their terms, "we want to consider poverty and educational level and illegitimacy and many other social indices and pathologies as an effect rather

than a cause—in social science terminology as a dependent not an independent variable" (pp. 129-130).

The extensive discussion of intelligence and its relationship to the presence or absence of financial, occupational, or educational success leads to one of the most controversial issues in *The Bell Curve*. If it is intelligence rather than class or race that is most influential in determining one's job, socioeconomic status, or general success in life, then the question of the source of IQ looms large. This is the reason for so much of the volatile and heated controversy that surrounds this book and its general theme.

First a note of clarification. The authors of *The Bell Curve* are not arguing that IQ is primarily genetic and therefore minimally or not at all vulnerable to modification. They state clearly and without equivocation:

> If the reader is now convinced that either the genetic or environmental explanation has won out to the exclusion of the other, we have not done a sufficiently good job of presenting one side or the other. It seems highly likely to us that both genes and the environment have something to do with racial differences. What might the mix be? We are resolutely agnostic on that issue; as far as we can determine, the evidence does not yet justify an estimate. (p. 311)

To add to this, a box on page 410 describes the significant influence of the environment on human development and intelligence. It uses the example of a feral child and concludes with this sentence: "If the ordinary human environment is so essential for bestowing human intelligence, we should be able to create extraordinary environments to raise it further." The authors observe that the relationship of IQ to intelligence is not well understood.

> Even so, the instability of test scores across generations should caution against taking the current ethnic differences as etched in stone. There are things we do not yet understand about the relation between IQ and intelligence, which may be relevant for comparisons not just across times but also across cultures and races. (p. 309)

But the potential optimism present in this part of the book is dashed in other sections when the authors state: "An individual's realized intelligence, no matter whether realized through genes or the environment, is not very malleable" (p. 314); "For many people there is nothing they can learn that will repay the cost of teaching" (p. 520).

Could the situation be that bad? Are the options so limited in number and extent that some groups of individuals should be written off? Are there any successful programs anywhere or is the only possible conclusion that:

> Taken together, the story of attempts to raise intelligence is one of high hopes, flamboyant claims and disappointing results. For the foreseeable future, the problems of low cognitive ability are not going to be solved by outside intervention to make children smarter. (p. 389)

Two points need to be made here. One is that it may be time to back off, cut a little slack, or direct a little "benign neglect" to the significance and value of IQ numbers and instead look for concrete, specific ways of helping students who are disadvantaged, from whatever source, to simply do better in schools. (*The Bell Curve* is a substantial and informative book, but is it possible it is too concerned with IQ scores? If one had a choice between having a high IQ score

or doing well in school, which would most people choose?) There are programs that are effective, in some cases highly effective, in helping the disadvantaged to do well in school. Following is a description of a school in the middle of the Chicago ghetto. It is full of black and generally poor children whose academic performance, if *The Bell Curve* is correct, should be at the lowest levels.

> As the patience of whites for other whites wears thin, the black inner city will simultaneously be getting worse rather than better. Various scholars, led by William Julius Wilson, have described the out migration of the ablest blacks that has left the inner city without its former leaders and role models. Given a mean black IQ of about 85 and the link between socioeconomic status and IQ in ethnic populations, the implication is that the black inner city has a population with a mean IQ somewhere in the low 80s at best, with a correspondingly small tail in the above-average range. (p. 522).

Somehow, this school that has a very clear, highly disciplined, and structured program with very high expectations of its students has "achieved honors as an academic institution above the national norms in all disciplines."

> In the Chicago ghetto today the only institutions with a record of consistently getting people out of the underclass are the parochial schools. They pay their teachers much less than what public-school teachers are paid, but they can screen their applicants, their principals can hire and fire, and they can and do impose many rules on both the students and their parents. (Ghetto public "magnet" schools that are allowed to screen are also successful.) Father George Clements, the pastor of the Holy Angels Catholic Church, describes the regimen at its elementary school this way: "We have achieved honors as an academic institution above the national norm in all disciplines. We bear down hard on basics. Hard work, sacrifice, dedication. A twelve-month school year. An eight-hour day. You can't leave the campus. Total silence in the lunchroom and throughout the building. Expulsion for graffiti. Very heavy emphasis on moral pride. The parents must come every month and pick up the report card and talk to the teacher, or we kick out the kid. They must come to the PTA every month. They must sign every night's homework in every subject. They must come to Mass on Sundays. They must take a required course on the Catholic faith. The kids wear uniforms, which are required to be clean, pressed, no holes. We have a waiting list of over a thousand, and the more we bear down, the longer the list gets. (Lehman, 1986, p. 69)

But Holy Angels is not the only Catholic school in the US that appears to be effective in educating those at the lower levels of American society. James Coleman, the American sociologist compared the effectiveness of public and parochial schools in the US. He found that it is useful to note that Catholic elementary schools in the US get no public funding, pay their teachers much less than public school teachers get paid, generally have little or no technical support, and spend about half to one third less on each of their students than their public counterparts. Despite all this, they do a much better job of educating their students, many of whom come from the lower echelons of American society than the public schools do.

In their book *Public and Private Schools*, Coleman and Hoffer (1987) compared Catholic and public schools over a three-year span. They found that poorly funded and equipped Catholic schools did a much better job of educat-

ing their students than the public schools did. In that three-year span, the Catholic school students gained almost one full grade equivalent in verbal and mathematical abilities.

A third example that is even more dramatic than the previous two may help. On page 444 of *The Bell Curve* this observation appears:

> Our proposal will sound, and is, elitist, but only in the sense that, after exposing students to the best the world's intellectual heritage has to offer and challenging them to achieve whatever level of excellence they are capable of, just a minority of students has the potential to become "an educated person" as we are using the term. It is not within everyone's ability to understand the world's intellectual heritage at the same level, any more than everyone who enters college can expect to be a theoretical physicist by trying hard enough. At every stage of learning, some people reach their limits. This is not a controversial statement when it applies to the highest levels of learning. Readers who kept taking mathematics as long as they could stand it know that at some point they hit the wall, and studying hard was no longer enough.

But the film *Stand and Deliver* may be instructive. It describes the teaching approach, work ethic, and perhaps most importantly the high expectations of Jaime Escalante for his lower-class Latino students. By dint of hard work, perseverance, diligence, discipline, and a profound commitment to their academic development, Escalante taught his students enough advanced math so that they could pass the Educational Testing Services (ETS) calculus exam. As the film makes clear (and as actually happened) the representatives of ETS did not believe that lower-class Latino high school students could learn math and especially calculus. Like Herrnstein and Murray, they assumed that you can't get blood (i.e., passing grades on an ETS calculus exam) out of stone (i.e., lower-class Latino students). At the end of the film, there is an indication that more and more lower-class Chicano students from that school took and passed the calculus exam. In an interview, Edward Olmos, who played the part of Escalante in the movie, noted that of the original 18 students who took and passed the calculus exam, 17 went to and finished college. Perhaps Mill was correct when he suggested that "one should always aim high and when you achieve that you won't achieve only that."

It is recognized that none of these examples has control groups or is capable of statistical analysis, and they are certainly not very empirical in their data or results. But could that not be positive? They give a sense of what is possible, not what is measurable.

These effective programs demonstrate that the problems that characterize North American education today are not going to be resolved, or perhaps even diminished, by *The Bell Curve* or any book like it. Perhaps Tyrell's (Schumacher, 1973) distinction of convergent and divergent type problems will be helpful. Most serious human problems and especially those associated with education are not of the convergent variety. A convergent problem is one that is open to logical, mathematical, or purely rational solutions. Math problems and many scientific questions are of this nature. Divergent problems are those that emanate from and are expressive of human relationships in all their myriad possibilities in human communities. These difficult issues are seldom open to resolution by mathematical or logical procedures. *The Bell Curve* presents all its

findings in a convergent mathematical format. Social science data like those found in *The Bell Curve* can provide useful information, but are usually not complete or clear enough to indicate clear and specific solutions. In opposition to this, divergent problems generally express issues for which there are no quick or easy or uniform answers. They encourage or demand that we go beyond what we clearly know or can clearly understand into realms where individual or social values are the determining factors.

Convergent problems are much easier to deal with than divergent ones. They often have distinct, clear answers and are readily and effectively expressed through charts, graphs, numbers, and various statistical or mathematical techniques. *The Bell Curve* presents its material as if the question of IQ and its relationship to class or race or social pathology were simply a matter of numbers and charts and lines on graphs. But there are distinct and highly significant social norms involved in how we value, express, and respond to variables like cognitive ability or intelligence. We may need a book that analyzes these social values as much as we need a book like *The Bell Curve* that seems to avoid them. Our most significant and complex social problems and especially the numerous problems associated with the process of education are not going to be resolved or even approached by a recipe-like process of putting in two egg charts, a tablespoon of statistical analyses, and then sifting in three graphs in the hope that *an* or *the* "answer" will be discovered after 45 minutes in the social oven. If we begin to value the statistical findings, graphs, and charts that constitute so much of *The Bell Curve* over the individual students in our schools or citizens in our cities then any chance of accurately envisioning, much less solving, our social problems may be lost. Keyes (1975) said this cogently in his novel *Flowers for Algernon*:

> "Don't misunderstand me," I said. "Intelligence is one of the greatest human gifts. But all too often a search for knowledge drives out the search for love. This is something else I've discovered for myself very recently. I present it to you as a hypothesis: Intelligence without the ability to give and receive affection leads to mental and moral breakdown, to neurosis, and possibly even psychosis. And I say that the mind absorbed in and involved in itself as a self centered end and to the exclusion of human relationships can only lead to violence and pain." (pp. 173-174)

The Bell Curve is a useful, perhaps necessary, book, but it has to be seen and understood in reference to the values of North American society. To view it outside that framework could be not just invalid, but harmful.

References

Cavalli-Sforza, L.L., Menozzi, P., & Piazza, A. (1993). *The history and geography of human genes.* Princeton, NJ: Princeton University Press.

Cohen, M. (Ed.). (1961). *The philosophy of John Stuart Mill.* New York: Random House.

Coleman, J., & Hoffer, T. (1987). *Public and private schools: The impact of community.* New York: Basic Books.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Keyes, D. (1975). *Flowers for Algernon.* Toronto: Bantam.

Lehman, N. (1986, June). The origins of the underclass. *Atlantic Monthly, 258*(1), p. 69.

Rosen, J., & Lane, C. (1994, October 31). Neo-nazis. *New Republic, 4*(18), p. 14.

Ryan, A. (1994, November 17). Apocalypse now? *New York Review of Books, XLI*(19), p. 11.

Schumacher, E.F. (1973). *Small is beautiful.* London: Blond and Briggs.

The Alberta Journal of Educational Research *Vol. XLI, No. 3, September 1995, 367-376*

David F. Robitaille

and

Edward C. Robeck
University of British Columbia

Distorted Vision: Education as Seen Through *The Bell Curve*

In the world according to *The Bell Curve* human lives are played out within boundaries determined by IQ and its correlates. The work of educators fits within that area, confined by students' "intellectual capacity" and configured by the interminable force that capacity exerts on each student's movement in society. In that world teaching begins by recognizing the trajectory nature has set for students as a result of their intellectual endowments, and progresses through a kind of intellectual and social streamlining of students—a shaping of their capacities and desires so they can be carried along, quickly and efficiently, to their proper and inevitable "place" in society. Herrnstein and Murray (1994) suggest this world as the focus of policy, a vision of the future that educators can use to guide their work. Although the data that they rely on to support their arguments for this vision can be (and have been) questioned in a variety of ways (Gould, 1994), we agree with those who suggest that this book and the vision it promotes may nonetheless become a "subtext for virtually every policy decision made concerning race, class, and social welfare" for at least the near future (Scott, 1994, p. 59). It is important to consider the vision Herrnstein and Murray present, even if its empirical support is fallacious, for the direction in which it might lead us as educators. We argue that the vision they offer would distort the purposes and practice of education by its selective attention to a single human attribute that they consider in relation to a limited set of educational factors and highly constrained set of educational goals.

Herrnstein and Murray would have us believe that social privilege and injustice are nothing other than articulations of innate propensities. These propensities are directly related to a variety of characteristics that are recognized as the basis of social class, which in turn affects much of social life. "Social class remains the vehicle of social life," they tell us, "but intelligence now pulls the train" (p. 25). It is an interesting metaphor, and similar to one that Walzer (1983) uses to describe the influence that the valuing of a particular attribute to the exclusion of others has in matters of social inequality and injustice:

David Robitaille is a professor of mathematics education and head of the Department of Curriculum Studies. His research focuses on evaluation of students' learning and international comparisons.

Ed Robeck is a doctoral student in science education and has recently accepted a position as assistant professor of science education at Hastings College in Nebraska. His research interests focus on ethics related to curriculum.

> Physical strength, familial reputation, religious or political office, landed wealth, capital, technical knowledge: each of these, in different historical periods, has been dominant; and each of them has been monopolized by some group of men and women. And then all good things come to those who have the one best thing. Possess that one, and the others come in train. (p. 11)

The Bell Curve is essentially about what the authors hold to be the "one best thing," intelligence. Other aspects of social life—such as school achievement, accomplishments in the work place, a stable family life—are the direct and indirect products of this fundamental capacity that they present as predominating all others.

Herrnstein and Murray's vision is based on a simplistic version of distributive social justice. Inequality is the result of the unequal distribution of intelligence, the universal and singular currency of social life. The dominance of intelligence provides a social monopoly to those who possess it to the greatest degree, that is, those that have the highest IQs. In *The Bell Curve* the monopoly of intelligence is compounded by Herrnstein and Murray's claim that it is both pivotal to success in all areas and genetically determined. They conflate and naturalize social supremacy of the well bred—aristocracy—with social supremacy of the highly talented—meritocracy—into a system of inescapable domination, which is in many respects outside of the control of humanity. Social inequity is, in Herrnstein and Murray's formulation, the result of natural differences in intelligence.

An important factor in Walzer's (1983) discussion of social justice is his assertion that there is no single standard for what will be valued in society. A given attribute attains a position of dominance as a matter of what has come to be appreciated and rewarded by a group. It is not due to the intrinsic worth of the attribute, but to the significance assigned to it. Artistic talents, physical strength, or emotional sensitivity could be dominant if a group came to value them over all others. Herrnstein and Murray apparently recognize this, agreeing that "the concept of intelligence has taken on a much higher place in the pantheon of human virtues than it deserves" (p. 21). At the same time they argue for intelligence as the singular standard by which to judge social life and set social policy. So, for example, although they are scrupulous in their disavowal of social engineering, they complain that current social policies encourage the "wrong women" (p. 548) to have children, a judgment they base on the mean IQ and other statistically assigned attributes of the groups with which they have identified the women. *The Bell Curve* is in effect an argument for an ideology in which social power is primarily and rightfully ordained first and foremost by intelligence.

With this focus on intelligence, attention turns away from other aspects of human life. In lived experience attributes such as caring, sensitivities to justice, creativity, intelligence, and others interact so that each affects what is done, with none of them predominating exclusively. Certainly human life is not just a matter of applying the singular attribute that is appropriate in a given situation. Rather than recognize this, Herrnstein and Murray strive to extend the dominance of IQ by presenting data they insist demonstrate its pervasive influence and with arguments to the effect that those who are most intelligent are also most valuable to society. The dominance of intelligence becomes

monopoly, and even tyranny, as it invades even the most personal aspects of life (Walzer, 1983, p. 18). Intellect is portrayed as rightfully dominating all decisions, including those involving childrearing, marriage, and where we might live. The pervasiveness they claim for the role of IQ is all the more troubling with the realization that whatever it is that is being measured by the tests Herrnstein and Murray are putting such faith in is certainly narrower than the attributes it will be taken to represent. That is, of course, assuming that the tests measure anything at all.

This concern with the distribution of a single characteristic carries over into their discussion of education as well. In general, they argue that the distribution of intelligence should influence the distribution of educational resources—students to whom nature has given more intelligence should be given more attention, more funding, and more challenging standards than they are now. Whatever a student might have to offer other than what shows up on IQ tests would be either neglected or granted only subaltern status as intelligence becomes the characteristic that defines students' position in school. Witherell (1991) points out the danger for educators of becoming too sure of what is important to know about students as she warns that

> When conceptions of the person, self, and community are not continually called into question in professional practice, reified and reductionist concepts emerge as common practice, creating new forms of "common sense" in the profession. Examples of such practices include excessive reliance on normative measures of aptitude, intelligence, psychopathology, values, or developmental states for educational or psychological assessment. (p. 84)

Considerations of particular aspects of human thought and experience in isolation are often useful for describing how they might mitigate learning, understanding, decisions, and other processes. However, the focus on any particular attribute or set of attributes has to be considered strategic and provisional, with the attributes being primarily artifacts of the assessments themselves. "Too often," Witherell continues, "diagnostic instruments take the place of the attention and dialogue that the practitioner needs in order to understand individuals in the context of their personal and cultural environments" (p. 84).

The need to address such contexts has been pointed out repeatedly in education, and in particular with respect to the inadequacy of the simpler forms of distributive justice that attend only to who has or is getting how much of something (Troyna & Vincent, 1995). Connell (1992) has stated that, "education is a social process in which the 'how much' cannot be separated from the 'what'" (p. 5). That is, the scope of education cannot be meaningfully discussed without also attending to its content. The explicit discussion that Herrnstein and Murray offer regarding the "how much" of education, therefore, also implicitly conveys messages as to the "what" that they would promote. They recall "the old days" when

> Standards would have been raised if students were to read a larger number of the Great Books (no one would have had much quarrel about what they were) or if students were required to write longer term papers, subject to stricter grading on argumentation and documentation. (p. 432)

What is denied by the parenthetical comment in the remark above is the importance of the discussions that must be undertaken regarding the "what" of the curriculum, including what should be read. For Herrnstein and Murray, the struggle that began in the 1960s to have the curriculum reflect the multicultural character of our society, and specifically the challenge that this poses for a curriculum based on an elite European tradition, served to "corrupt" (p. 432) educational standards. Perhaps they are forgetting the arguments that have been made regarding the relationship of knowledge and power in society. Perhaps they are forgetting how that relationship has been shown to extend into schools through the "deep imbrication of traditional, canonical school knowledge in the legitimation of authority and inequality in society" (McCarthy, 1993, pp. 289-290). But such critiques, which are still being made (Bishop, 1990; Castenell & Pinar, 1993; Joseph, 1987; see also Pinar, Reynolds, Slattery, & Taubman, 1995, section III), must be remembered in any discussion of education, for they have a great deal to say about the legacies of imperialism, the ongoing struggles to understand the role of curricula in racism and sexism, and the inequities that result from favoring certain forms of knowledge while overlooking or deriding others. Allowing such critiques to be forgotten, or more likely actively neglected, reinforces the subjugation and marginalization of students whose personal and/or cultural contexts differ from what is assumed to be proper curricular knowledge. No approach to education can provide hope for social justice if it does not recognize the history of struggle over curricular knowledge and the need to continue these struggles continuously as a hedge against the inclination to universalize dominant knowledge.

The injustices that follow from neglect of the importance of context are also supported by the structure of the arguments presented by Herrnstein and Murray, particularly with respect to the implications that such neglect would have for policies in education and other areas. The book discusses education and a range of other social issues in terms of generalizations such as group means. Herrnstein and Murray indicate that they appreciate the limits of such generalizations, and early in the book readers are told that "*Measures of intelligence have reliable statistical relationships with important social phenomena, but they are a limited tool for deciding what to make of any given individual*" (p. 21 original emphasis) and that "this thing we know as IQ is important but not a synonym for human excellence" (p. 21).

However, in everyday interactions people are all too often identified as members of groups before they are identified as individuals, and it is here that stereotypes can have a major impact on relationships. Posner (1981) presents the case that assumptions about individuals are often made by invoking the expediency of assuming a member of a group shares traits associated with the group generally:

> To the extent that race or some attribute similarly difficult to conceal (sex, accent) is positively correlated with undesired characteristics or negatively correlated with desired characteristics, it is rational for people to use the attribute as a proxy for the underlying characteristics with which it is correlated. (p. 362)

This is what is referred to as discrimination based on "information cost." It is premised on the idea that gaining the information needed to assess individuals

with respect to the assumed traits is too costly (or too risky) in terms of time, energy, finances, and so on. To economize on information costs, group trends are taken as sufficient to represent the individual. When this is bolstered by statistical correlations such as those Herrnstein and Murray present, it provides the basis of, for example, the "'probabilistic turn' in which racial identity comes to be rewritten as a statistical property" (Gates, 1992, p. 331). With respect to identifications of "race" and other attributes, stereotypes associated with groups are themselves a matter of selective attention and prejudiced representations, but in *The Bell Curve* they take on renewed status and significance.

Invoking a probabilistic turn in several areas at once, *The Bell Curve* shifts from early discussions of the "cognitively disadvantaged" to "the disadvantaged" to definitions of the disadvantaged that include all children served by a range of federal programs in the United States:

> The programs we designated as for the disadvantaged were the [Elementary and Secondary Education Act of 1965 (ESEA)] Title I basic and concentration grants, Even Start, the programs for migratory children, handicapped children, neglected and delinquent children, the rural technical assistance centers, the state block grants, inexpensive book distributions, the Ellender fellowships, emergency immigrant grant education, the [ESEA] Title V (drug and alcohol abuse) state grants, national programs and emergency grants, [ESEA] Title VI dropout, and bilingual program grants. (p. 753)

In *The Bell Curve*, then, social characteristics that correlate with achievement become proxies for academic potential. Are we to suppose, for example, that because the immigrant population is "probably somewhat below the native-born average" (p. 364) all immigrant children should be assumed to be cognitively disadvantaged? Similarly, are all bilingual children, migratory children, handicapped children, and others to be considered at the low end of the cognitive distribution, rather than disadvantaged due to lack of dominant cultural experiences, access to resources, or other limitations as was intended by the original framing of the ESEA (MacKintosh, Gore, & Lewis, 1965)? In the original framing, the ESEA was intended to assure that students' opportunities not be constrained by economic, social, personal, and other factors that were not intrinsic to the student. Alternatively, relying on group trends, which seems to be part of the vision Herrnstein and Murray offer educators, could end up truncating students' opportunities because of the limitations they are assumed to have as selected features that are statistically associated with their demographic backgrounds, taken to indicate their personal characteristics.

As in the discussion of context above, the reductionism of such an approach threatens the integrity of education. Posner (1981) points out that making decisions based on information costs leads to lost opportunities for "valuable associations" with individuals who are treated only as group members (p. 362). For educators, the associations that would be precluded in efforts to economize on information costs are the very associations with students required for effective pedagogical relationships in which students' needs, abilities, interests, and desires are known and responded to by educators. To forgo such associations is to assume beforehand that some students, who can be identified largely by demographic factors, will simply not make it in school or society to become

anything but subordinate, whereas others have a social profile that shows that they are obviously going to excel, succeed, and lead. It is reasonable, if one believes *The Bell Curve*, and it is efficient; it is also patently unjust.

The concerns with the support the book offers for approaches that economize on information costs are made all the more pronounced when one considers that there are already significant effects on children due to various social biases in schools. A powerful example of this can be seen in the results of the Second International Mathematics and Science Study (SIMS), which Herrnstein and Murphy cite along with others to support their contention that gifted students are not being led to develop their full potential in schools (p. 417). Taking a trivial "horse race" view that is common in many approaches to the results of international studies, Herrnstein and Murray initiate a discussion relating achievement to their concern with the distribution of educational resources. They begin by pointing out that the US typically scores in the middle or bottom of international comparisons. They use these results to bolster their case that gifted students need more than they are getting from educators.

In outlining what might be done, however, they ignore the influences of major sources of variability on students' achievement scores. Given their focus on the importance of IQ in their analysis and recommendations, one might expect the greatest source of variation in students' achievement to be differences in students' ability, that is, differences between students would explain more of the variance in students' scores than any other factor. In fact, as Figure 1 shows, in the US between-classroom differences accounted for slightly more (47%) of the variance in grade 8 students' achievement in SIMS than did differences between students (46%); the remainder is between-school variance (Schmidt, Wolfe, & Kifer, 1992). In Japan, on the other hand, 91% of the variance in students' achievement was attributable to differences between students and only 9% to differences between schools. Between-classroom variation in Japan was negligible. What this means is that in the US the classroom or course in which a student was enrolled was at least as important—in terms of accounting for students' achievement—as factors personal to the students: IQ, motivation, socioeconomic status, and so on. Figure 1 shows the relative strength of the variance components for Japan and the US.

The size of the between-classroom variance component for the US is largely a function of pedagogical practices such as streaming or tracking of students into different courses and curricula. It is useful to consider, then, how students come to be placed in different mathematics classes.

In general, throughout the US, the biggest difference in mathematics classes at grade 8 has to do with which students take algebra, a course that at that level is intended for students with high ability in mathematics. Decisions as to which students will take algebra in grade 8 are made locally, at the school level, and the criteria used to make the selection vary across schools. Results from SIMS show that this decision making process is weak in that many students placed in nonalgebra courses actually have higher achievement scores than many algebra students. Only about half of the top 10% of the US students on the SIMS, and a third of the top 25%, were taking algebra, indicating that many errors were made in determining which students were the most capable (Kifer, 1992).



Figure 1. Variance components for school, classroom, and student in SIMS (adapted from Schmidt, Wolfe, & Kifer, 1992).

Taking a closer look at who the students were that were taking algebra, some troubling indications emerge. Of the highest achieving 25% of US students in SIMS, 91% were white. So one might reasonably expect that about this same proportion of the students taking algebra would be white. As Figure 2 shows, the actual number is fairly close, 92%. The expected and actual proportions are also quite close for the number of girls enrolled: 49% expected and 50% actual. When father's occupation is considered (which was used as a proxy for SES) there is more disparity: 32% would be expected to come from families in which the father was in a professional occupation, whereas the actual number is 40%. When the students who scored in the second 25% on SIMS are considered, the divergence between expected and actual numbers of students enrolled becomes pronounced. The results show clearly that white students, girls, and students whose fathers were professionals were disproportionately likely to be selected for grade 8 algebra (Kifer, 1992).

Each of these three attributes should be expected to be irrelevant to selection for algebra, unless they were being used in an atmosphere where "race," gender, or family background were assumed to be related in a causal way to ability or academic achievement. *The Bell Curve* provides evidence construed in a way that supports such assumptions. This has serious implications because of the way the high school curriculum in the US is structured. Entry into high-status courses such as calculus is the culmination of a sequence of courses taken over several years. Therefore, the opportunity to take those high-status courses is almost always restricted to those students who studied algebra in grade 8. Thus the classification decision in grade 8 has important consequences for individual students' opportunities.

These results suggest that the US educational system is undermining the achievement of many students, by the meritocratic practice of selecting students for homogeneous grouping based on achievement. In more egalitarian systems, such as Japan where students of all abilities learn in the same classroom at the grade 8 level, the achievement of all students is enhanced.

As is often the case in education, the implications of the SIMS results are not apparent from a trivial look at the data. Questions in education cannot be reduced simply to who scores higher on a test. Herrnstein and Murray's

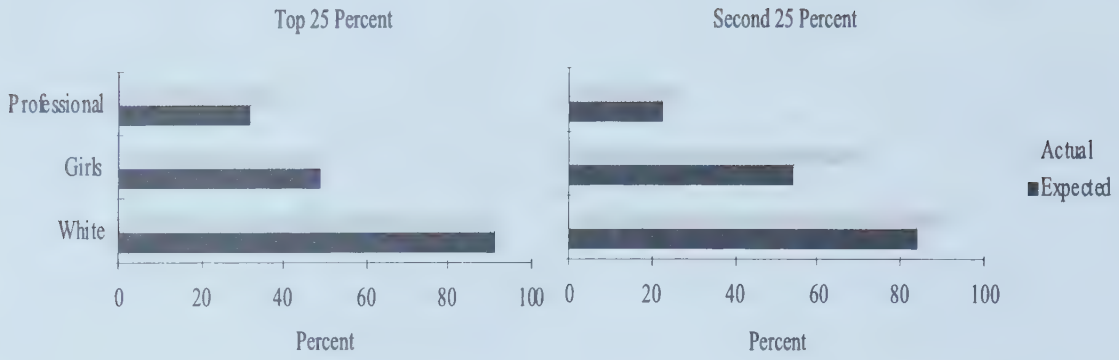


Figure 2. Participation rates in grade 8 algebra in the United States from SIMS (Adapted from Schmidt et al., 1992).

reliance on such indicators without reference to more subtle factors may well lead to an intensification of the same meritocratic practices such as homogeneous grouping that seem to have contributed to the results they use as evidence for the problems they would have educators address.

Also, although it may not concern the authors of *The Bell Curve* in their exclusive attention to achievement, other studies have shown that identifying students as gifted for the purposes of special programming seems to have the effect of inhibiting creativity and contributes to a variety of social and emotional problems experienced by the students (Freeman, 1994a, 1994b). There are certainly other studies that would offer still more considerations regarding how to proceed. The point is that many factors, including the contexts, contents, and practices of education must be considered in determining policy, and with reference to more than a single scale such as IQ.

A final point must be made with respect to Herrnstein and Murray's deliberate attempts to divert attention from some of the most serious implications of their work, namely, the realization that their recommendations are predicated on the perpetuation and even the intensification of social injustice. In addition to providing data and arguments that can be applied to various sites at which biases can operate such as selection for high-status classes, *The Bell Curve* maintains a position throughout that inequities, even those with dire consequences, are inevitable, and educators should make the best of them. In fact, the authors are quite explicit about telling readers to disregard the inequities that are quite obviously being promoted as based on natural laws. Besides, to ask questions as to the effects of focusing on one segment of the student population is irrelevant, or so readers are told, because "the answers to such questions have nothing to do with social justice but much to do with the welfare of the nation, including the ultimate welfare of the disadvantaged" (p. 442). They go on to say:

> To be intellectually gifted is indeed a gift. Nobody "deserves" it. The monetary and social rewards that accrue to being intellectually gifted are growing all the time, for reasons that are easily condemned as being unfair. Never mind, we are saying. The gifted youngsters are important not because they are more virtuous or deserving but because our society's future depends on them. (p. 442)

In this twisted rhetoric, increased injustice for those who are already treated unjustly is deemed acceptable, and probably becomes necessary for later justice. This deferred justice, a vision of which is apparently not even imaginable, will come at the hands of those who are already privileged, because they will have the opportunity to find a way out of unjust social conditions. The outrage the reader might feel at inequitable societal patterns is to be assuaged by advice offered about what to do instead of trying to dismantle privilege. They state, for example, that

> Most gifted students are going to grow up segregated from the rest of society no matter what. They will then go to the elite colleges no matter what, move to successful careers no matter what, and eventually lead the institutions of the country no matter what. Therefore, the nation had better do its damnedest to make them as *wise* as it can. If they cannot grow up knowing how the rest of the world lives, they can at least grow up with a proper humility about their capacity to reinvent the world de novo and thoughtfully aware of their intellectual, cultural, and ethical heritage. They should be taught their responsibilities as citizens of a broader society. (p. 443, original emphasis)

Embedded in this view is the assumption that society's future does not depend on the rest of the students, those who are not so lucky as to have been gifted. The future, at least the one that would take society in what the book projects as a positive direction, certainly does not depend on students at the lower end of the cognitive distribution, regardless of what else they might have to offer society. What the book invokes, then, is a form of social triage, or placing students in groups that lie on a continuum with respect to the likelihood of their academic success and approaching their instruction accordingly (Books, 1992). At one end are those who have the dominant characteristic, intelligence, and are most likely to provide benefits to society in important and unique ways. In the middle are students who are necessary, but are eminently replaceable in their insuperable averageness. At the other end are students who are expendable with respect to what they might offer society. In *The Bell Curve* it is this triage that determines the dispersal of educational resources as concern for individual students is overwhelmed by a faith in indeterminate collective progress.

Conclusion

Throughout *The Bell Curve* a single human attribute is privileged. The book presents an argument that most human differences can be traced to it and almost all social injustices explained by it. The social struggles that have invigorated the work of teachers over the past several decades, as well as presenting them with their most important challenges, are neglected or are presented as forces that have corrupted and degraded education. To suggest policy with such a limited focus can only debase the goals of education and pervert educators' relationships with students.

Much work has already been done toward understanding the role of education in aspects of students' lives that intersect with intellect, but also involve awareness that goes beyond what intellect alone can offer. Several authors have addressed the ethical and moral aspects of both curriculum and instruction (Goodlad, Soder, & Sirotnik, 1990; Jackson, Boostrom, and Hansen, 1993;

Witherell and Noddings, 1991). What we see in these areas is a more complex vision of social justice than that invoked by Herrnstein and Murray. In such a vision, society would be sensitive to and value a variety of human attributes. Social interactions would involve reliance on constantly shifting asymmetries of not only intelligence, but also creativity, empathy, aesthetic appreciation, caring, and other aspects of human life. Attending to a variety of human attributes while also respecting the effects of students' social and personal contexts would offer a vision of education that could enhance social justice not only in some distant undreamt of future, but also in the day-to-day practices of teaching in the present.

References

Bishop, A.J. (1990). Western mathematics: The secret weapon of cultural imperialism. *Race and Class, 32*(2), 51-65.

Books, S. (1992). Literary journalism as educational criticism: A discourse on triage. *Holistic Education Review, 5*(3), 41-51.

Castenell, Jr., L.A., & Pinar, W.F. (Eds.). (1993). *Understanding curriculum as racial text: Representations of identity and difference in education*. Albany, NY: SUNY Press.

Connell, R.W. (1992, January). *Citizenship, social justice, and curriculum*. Paper presented at the International Conference on Sociology of Education, Westhill.

Freeman, J. (1994a). Gifted school performance and creativity. *Roeper Review, 17*(1), 15-19.

Freeman, J. (1994b). Some emotional aspects of being gifted. *Journal for the Education of the Gifted, 17*(2), 180-197.

Gates, H.L. (1992). Statistical stigmata. In D. Cornell, M. Rosenfeld, & D.G. Carlson (Eds.). *Deconstruction and the possibility of justice* (pp. 330-345). New York: Routledge.

Goodlad, J.I., Soder, R., Sirotnick, K.A. (Eds.). (1990). *The moral dimensions of teaching*. San Francisco, CA: Jossey-Bass.

Gould, S.J. (1994). Curveball. *The New Yorker, 70*(39), pp. 139-149.

Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life*. New York: Free Press.

Jackson, P.W., Boostrom, R.E., & Hansen, D.T. (1993). *The moral life of schools*. San Francisco, CA: Jossey-Bass.

Joseph, G.G. (1987). Foundations of Eurocentrism in mathematics. *Race and Class, 28*(3), 13-28.

Kifer, E. (1992). Opportunities, talents and participation. In L. Burstein (Ed.), *The IEA Study of Mathematics: Student growth and classroom processes* (pp. 279-307). Oxford: Pergamon.

MacKintosh, H.K., Gore, L., & Lewis, G.M. (1965). *Educating disadvantaged children in the primary years: Kindergarten through grade 3*. Washington, DC: Bureau of Educational Research and Development.

McCarthy, C. (1993). After the canon: Knowledge and ideological representation in the multicultural discourse on curriculum reform. In C. McCarthy & W. Crichlow (Eds.), *Race, identity, and representation in education* (pp. 289-505). New York: Routledge.

Pinar, W.F., Reynolds, W.M., Slattery, P., & Taubman, P.M. (Eds.). (1995). *Understanding curriculum*. New York: Peter Lang.

Posner, R.A. (1981). *The economics of justice*. Cambridge, MA: Harvard University Press.

Scott, D.M. (1994). Cognitive conceit: A review of *The Bell Curve*. *Social Policy, 20*(2), 50-59.

Schmidt, W.H., Wolfe, R.G., & Kifer, E. (1992). The identification and description of student growth in mathematics achievement. In L. Burstein (Ed.), *The IEA study of mathematics: Student growth and classroom processes* (pp. 59-99). Oxford: Pergamon.

Troyna, B., & Vincent C. (1995, April). *The discourses of social justice in education*. Paper presented at the American Educational Research Association, San Francisco.

Walzer, M. (1983). *Spheres of justice: A defense of pluralism and equality*. New York: Basic Books.

Witherell, C. (1991). The self in narrative. In C. Witherell & N. Noddings (Eds.), *Stories lives tell: Narrative and dialogue in education* (pp. 83-95). New York: Teachers College Press.

Witherell, C., & Noddings, N. (Eds.). (1991). *Stories lives tell: Narrative and dialogue in education*. New York: Teachers College Press.

This special issue of the *Alberta Journal of Educational Research*, Canadian Perspectives on *The Bell Curve*, is available at $17.50 per copy (includes postage and handling) from:

Office of the Dean, Faculty of Education
845 Education Centre South
University of Alberta
Edmonton, AB T6G 2G5
Canada

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes;* any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
845 Education Centre South
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1438



THE ALBERTA JOURNAL OF EDUCATIONAL RESEARCH

VOLUME XLI NUMBER 4 DECEMBER 1995

The Alberta Journal of Educational Research

Published in March, June, September, and December by the Faculty of Education, University of Alberta.

AJER is a quarterly journal devoted to the dissemination, criticism, interpretation, and encouragement of all forms of systematic enquiry into education and fields related to or associated with education.

Editor: Judy Cameron

Managing Editor: Naomi Stinson
Proofreader: Karen McFarlane
Translator: Ronald N. Boivin

Consulting Editors

Terry Belke
Mount Allison University

George Buck
University of Alberta

Ardra L. Cole
Ontario Institute for Studies in Education

John Connors
Canadian Union College

Samuel Deitz
Georgia State University

Sharon M. Haggerty
University of Western Ontario

Antoinette Oberg
University of Victoria

Ruth Rees
Queen's University

Heather Ryan
Massey University

Robert H. Short
University of Alberta

Kelleen Toohey
Simon Fraser University

Robert Wilson
Queen's University

Faculty Advisory Committee

John G. Paterson (Chair)
Joyce Edwards
Robert H. Short
Fern Snart
Ken Ward

AJER gratefully acknowledges support from the Social Sciences and Humanities Research Council of Canada, the Alberta Advisory Committee for Educational Studies, and the Research Grants Office of the University of Alberta

The subscription rate is $37.00 per year for individuals, $45.00 per year for institutions. Add $8.00 for delivery outside Canada. Single copies are $8.00 each. Subscriptions and sales in Canada will be charged 7% GST. Please make cheques payable to *The Alberta Journal of Educational Research*. Back issues are available; rates supplied on request. Claims for undelivered copies must be received within three months of publication.

Address all communications and manuscript submissions to:

The Alberta Journal of Educational Research
Office of the Dean
845 Education Centre South
University of Alberta
Edmonton, AB, Canada, T6G 2G5.
Fax: (403) 492-0236

Publications Mail Registration Number 1436

Faculty of Education
University of Alberta

The Alberta Journal of Educational Research

Volume XLI, Number 4, December 1995

Judy Cameron **377** From the Editor

Articles

Stephen P. Norris **378** Format Effects on Critical Thinking Test Performance

José L. da Costa **407** Teacher Collaboration: A Comparison of Four Strategies

Judith Golec
John Gartrell
William J. Sveinson **421** University Performance of Nonmatriculated Admissions

June Beynon
Kelleen Toohey **435** Access and Aspirations: Careers in Teaching as seen by Canadian University Students of Chinese and Punjabi-Sikh Ancestry

Marianne Gareau
Don Sawatzky **462** Parents and Schools Working Together: A Qualitative Study of Parent-School Collaboration

G. Patrick O'Neill **474** Authorship Patterns in AJER: Forty Years in the Making

R.S. Brown
J.O. Anderson
M.M. McGee **482** Manifestation of Latent Thetas: A Comparison of Field-Study and Positivistic Approaches to the Investigation of Self-Nurturance

ISSN 0002-4805

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 377-377*

From The Editor

This issue marks the end of my tenure as Editor. I would like to take this opportunity to thank the many individuals who made my experience as Editor rewarding. Several people deserve recognition for their efforts in producing a top journal. Thanks to all the authors who submitted papers and to the Editorial Board and reviewers who offered their critical and incisive comments. I would also like to express my gratitude to the staff in the Dean's Office who attended to aspects of finance, printing, and mailing, and to the *AJER* Faculty Advisory Committee for their support and encouragement. A special thanks to the managing editor Naomi Stinson for transforming manuscripts into the high quality journal that you are reading now.

It is my pleasure to introduce the new Editor Dr. Beth Young from Educational Policy Studies. Dr. Young brings a new perspective to *AJER*. Her research interests include gender issues in Canadian educational administration and leadership; her focus is on women's careers in education and related policy issues. On behalf of myself and all those involved in *AJER*, we welcome Beth to her new role.

Judy Cameron

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 378-406*

Stephen P. Norris
Memorial University of Newfoundland

Format Effects on Critical Thinking Test Performance

The influence of five format differences on the scores of 15 critical thinking measures was examined. The format differences were: whether the measures were in multiple-choice or constructed-response format; whether among the multiple-choice measures all items had a common set of response alternatives, or specific alternatives for each item or subset of items; whether or not examinees were requested to provide justifications for their answers; whether or not examinees were required to make judgments of credibility of information or of people; and whether or not the items were cast in the context of a narrative story. Using these format differences, a variety of construct models were hypothesized to explain correlations among scores on the measures. Confirmatory factor analysis was used to test the ability of each of the hypothesized models to account for the observed correlation matrix calculated from the performances of 172 senior high school students on all the measures. Only one model was able to pass all the tests of goodness of fit applied to the data, including the χ^2 test of perfect fit between the observed correlations and the correlations implied by the hypothesized model. That model postulated five factors: a common critical thinking factor, a narrative context factor, a nonnarrative context factor, a credibility judgment required factor, and a no credibility judgment required factor. One of the most significant conclusions from this finding is that much of the debate over the relative merits of multiple-choice and constructed-response critical thinking tests is ill founded: these two formats appear to be equivalent.

Cette étude examine l'influence de cinq différentes sortes ou formats de testing sur les scores de 15 moyens d'évaluation de la pensée critique. Les différences entre les sortes ou formats de tests étaient: si les réponses étaient du genre choix multiples ou réponses à développement; si parmi les réponses à choix multiples il existait un ensemble commun de réponses alternatives possibles, ou des alternatives spécifiques pour chaque item ou chaque sous groupes d'items; si on demandait ou ne demandait pas à chaque candidat(e) à l'examen de justifier ses réponses; si les candidat(e)s devaient ou ne devaient pas porter jugement sur la crédibilité de l'information ou des personnages présentés; et si les items étaient présentés ou pas dans le contexte d'une histoire au style narratif. C'est en utilisant ces différents formats et ces différentes sortes d'examens présentés qu'on a pu construire une variété de modèles pour expliquer la variation et la covariation des scores selon les différents moyens d'évaluation évalués. On utilisa l'analyse des facteurs de confirmation pour vérifier l'habilité de chaque moyen d'évaluation sur les facteurs hypothétiques pour expliquer la matrice de variation et de co-variation observées et calculées d'après les performances de 172 étudiant(e)s du niveau secondaire 2^{e} cycle sur les 15 moyens d'évaluation. Seulement un modèle semblait remplir les critères de satisfaction appliqués aux données y inclus le test χ^2 d'exactitude entre les corrélations observées et les corrélations sous-entendues par le modèle hyphothétique. Ce modèle a postulé cinq différents facteurs: un facteur de pensée critique commun, un facteur de contexte narratif, un facteur de contexte non-narratif, un facteur qui nécéssiterait un jugement de crédibilité, et un facteur qui requierrerait un jugement de non-crédibilité. Une des conclusions les plus significatives de cette recherche est que les mérites relatifs sur le débat des questions à choix multiples contre celles des questions à développement est en grande partie

Stephen Norris is a professor of educational research and philosophy of education in the Faculty of Education. His areas of research interest include the philosophy and policy of science education, philosophy of reading, reading science text, and critical thinking testing.

réponses mal fondé. Ces deux genres de questions, celles à choix multiples et celles à développement, semblent s'équivaloir.

Do differences in format affect what tests of critical thinking measure? This question generates much opinion and controversy, but there is insufficient evidence to answer it adequately. In this study the influence on scores of several format differences was examined for 15 measures of critical thinking.

The format difference that perhaps has caused most of the debate over critical thinking testing is between multiple-choice and constructed-response tests. Research on the relationship between multiple-choice and constructed-response tests began in the 1920s. According to Hogan (1981), who conducted an extensive review of studies up to 1981, the chief concern in the earliest investigations was the comparative reliability of the two formats. However, the concerns changed to include the question of equivalence. Hogan believed at the time of writing his report that regardless of the research "popular opinion on this question [was] rather well formulated and almost universally negative, i.e., the two types of items do *not* measure the same thing" (p. 2). This opinion is especially significant in the area of critical thinking testing. One argument used to support the opinion is that tests of subject matter that employ the multiple-choice format may be able to test effectively for knowledge recall, but it requires constructed-response tests to assess ability to think at higher levels about the content. Because critical thinking tests are by definition tests of higher level thinking, the conclusion is obvious: the term *multiple-choice critical thinking test* is an oxymoron. Hence the issue is salient in the domain of critical thinking as it is contended that as a matter of principle *nothing* in the domain is testable using multiple-choice tests. If this contention is true, then all critical thinking testing should follow the more time-consuming and expensive (by a factor of about 1,000, according to Ennis & Norris, 1990) route of constructed-response formats.

In his review, Hogan (1981) drew a series of conclusions based on three types of studies: direct correlation with correction for attenuation of multiple-choice and constructed-response tests to check for divergence of the correlation coefficient from unity; correlation of multiple-choice and constructed-response tests with a separate criterion measure to check which format yields the higher correlation; and treatment application to determine which type of measure is more sensitive to detection of a predicted change. His generalizations that are most relevant to this study are: (a) in most instances, multiple-choice and constructed-response measures were found to be equivalent or nearly so; and (b) multiple-choice tests tend to correlate more highly than constructed-response tests with external criteria.

In a more recent review Traub and MacRury (1990) drew different conclusions. Acknowledging that the differences are not well understood, they concluded that the two formats appear to measure somewhat different abilities, although any difference that exists is slight, if the constructed-response format requires answers no longer than a sentence or two for each item. Traub (1993) further argued that differences between the formats should not be expected if items require the manipulation of data and ideas. Barnett-Foster and Nagy (1995) found "the two formats are remarkably similar for first-year

university students in science and science-related programs of study" (p. 33). However, they cautioned that the conclusion should not be extended to tests of other types of reasoning strategies.

Thus there is sufficient uncertainty about the relationship between multiple-choice and constructed-response formats to warrant further research in the area. There is additional motivation to explore the issue using critical thinking tests, because they have not been studied previously with regard to this topic and because the opinion of many is that in principle critical thinking can be tested only with constructed-response formats. Even if the multiple-choice versus constructed-response difference makes no difference to critical thinking testing, it is still worthwhile exploring the influence of other differences in format.

Included among the 15 measures examined in this study were six in constructed-response format, and nine in multiple-choice format. In addition to variation along the multiple-choice/constructed-response format dimension, the 15 measures varied along several other dimensions that could affect scores. First, in the multiple-choice format four measures used a common set of alternatives for all items, and five measures used a specific set of alternatives for each item or each subset of items. Second, in the constructed-response format three measures requested examinees to justify their answers with reasons, and three measures did not request justification. Third, seven measures (4 in multiple-choice format and 3 in constructed-response format) asked examinees to judge the credibility of information presented; the other eight measures did not ask for such judgments. Finally, eight measures (5 multiple-choice and 3 constructed-response) embedded items in the context of story narratives; the other seven measures used either expository or descriptive portrayals of content for contextualizing items.

Based on these five format differences (multiple-choice vs. constructed-response; common-alternatives vs. specific-alternatives; justification-requested vs. justification-not-requested; credibility-judgment-required vs. credibility-judgment-not-required; narrative-context vs. nonnarrative-context), a variety of construct models were hypothesized to explain variation and covariation among scores. Confirmatory factor analysis was used to test the ability of each hypothesis to account for the observed correlation matrix among scores on the 15 measures.

More specifically, the construct models displayed in Table 1 were hypothesized. The models were based on combinations of the 15 factors listed in the left-hand column. The first group of six factors are simple factors because each refers only to one format characteristic. The second group of eight factors are compound factors because each refers to two format characteristics. Finally, a common critical thinking factor was used in building some models. Using various combinations of these 15 factors, 12 construct models were built: three each of three-factor, four-factor, five-factor, and six-factor models. An attempt was made to test each model under two assumptions: (a) with all factors constrained to be orthogonal, and (b) with at least some of the format factors correlated. In two instances, however, models with correlated factors were insufficiently constrained to allow estimation of the models to proceed.

Table 1
Hypothesized Construct Models to Explain Correlations Among 15 Measures

	Three-Factor Models			Four-Factor Models			Five-Factor Models			Six-Factor Models		
Factors	*3A*	*3B*	*3C*	*4A*	*4B*	*4C*	*5A*	*5B*	*5C*	*6A*	*6B*	*6C*
Simple factors												
multiple-choice (m-c)	X									X		
constructed-response (c-r)	X									X		
narrative context		X		X			X			X	X	
nonnarrative context		X		X			X			X	X	
credibility judgment required			X	X			X			X		X
credibility judgment not required			X	X			X			X		X
Compound factors												
m-c/common alternatives					X			X			X	
m-c/specific alternatives					X			X			X	
c-r/justification requested					X			X			X	
c-r/justification not requested					X			X			X	
m-c/narrative						X			X			X
m-c/nonnarrative						X			X			X
c-r/narrative						X			X			X
c-r/nonnarrative						X			X			X
Common critical thinking factor	X	X	X				X	X	X			

The three-factor models all included a common critical thinking factor. In each model the other two factors represented a format difference that divided the sample of measures into two subgroups. Two of the four-factor models (4B and 4C) divided the measures into multiple-choice and constructed-response categories, and in addition used other format factors to subdivide within each of these two main categories. Model 4A referred to format factors that cut across the multiple-choice and constructed-response categories. The five-factor models were constructed by adding a common critical thinking factor to each of the four-factor models. Finally, the six-factor models comprised combinations of the four-factor and three-factor models less the common critical thinking factor. It was not fruitful to combine the compound factors used in model 6B with the credibility judgment factors because the divisions m-c/ common-alternatives, m-c/specific-alternatives, c-r/justification-requested, and c-r/justification-not-requested were coextensive, respectively, with the categories m-c/credibility-judgment-required, m-c/credibility-judgment-not-required, c-r/credibility-judgment-required, and c-r/credibility-judgment-not-required.

Significance

The questions raised in this study have broad educational significance. Critical thinking has been advocated as an essential goal of education throughout this century (Black, 1946; Dewey, 1933; Ennis, 1962, 1981; Hullfish & Smith, 1968; Siegel, 1988; Smith, 1953). However, as Black (1946) said, "To be in a position to improve reasoning means to be in a position to distinguish good reasoning from bad" (p. 7).

Critical thinking is widely taken to be reasonable and reflective thinking that is focused on deciding what to believe or do (Norris & Ennis, 1989), a conception that has evolved from an earlier one by Ennis (1962), which took critical thinking to be the correct assessing of statements. So conceived, the critical thinker has both abilities and dispositions (Ennis, 1981; Perkins, Jay, & Tishman, 1993; Siegel, 1988). The abilities include the ability to maintain clarity, to provide and judge the basic support for actions and beliefs, and to make and judge inferences from basic information (Norris & Ennis, 1989). Dispositions to think critically are also needed, because people can have abilities they do not use. Critical thinkers are disposed, among other things, to seek reasons, consider alternatives, and withhold judgment when the evidence is not sufficient. This study focuses only on tests that are cast as measures of critical thinking abilities. Dispositions possibly play a role in determining performance on these tests, and possibly test format affects the size of that role. Nevertheless, disposition effects are not examined here.

As pointed out in several places (Bennett & Ward, 1993; Ennis & Norris, 1990), there has been a great deal of rhetoric about the adequacy of various formats for critical thinking testing. Some of this rhetoric has issued definitive conclusions about the relative merits of, for example, multiple-choice and constructed-response formats based on questionable claims about one format or the other such as: that scoring constructed-response tests is unreliable, that multiple-choice tests are more objective, that constructed-response tests give better evidence of thinking, that constructed-response tests assess productive and organizational skills, that multiple-choice tests cannot measure higher order thinking, and that constructed-response tests better resemble criterion

behaviors. Although these claims are often presupposed in recommendations for critical thinking testing, little evidence is available to support them. For instance, constructed-response formats are often proposed as a solution to perceived problems with multiple-choice formats on the grounds that constructed responses provide more direct and complete evidence on thinking than multiple-choice answers. However, this rationale relies on unsupported assumptions about the relative validity of different critical thinking test formats and unnecessary assumptions about how multiple-choice tests must be designed (Norris, 1992). Perhaps the two formats measure different constructs; perhaps they measure the same constructs. Perhaps well-designed multiple-choice tests could support the same inferences about critical thinking as are supported by constructed-response tests. Evidence on these and other issues regarding critical thinking testing is needed badly.

It is easy to agree with Resnick (1987) that critical thinking testing has not kept up with the demand for instruction in the area. There are no adequate answers to the questions raised above, despite their recognized importance (Arter & Salmon, 1987; Ennis, 1984; Ennis & Norris, 1990). Virtually all the evidence on critical thinking tests comes from factor analyses of superseded versions of the Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) and the Cornell Critical Thinking Tests (Ennis & Millman, 1985). Also many criticisms of such widely used critical thinking tests are based on inadequate evidence. For example, McPeck (1981) concluded only from first-order correlations of 0.55 to 0.75 between the Watson-Glaser test and IQ and reading tests that the Watson-Glaser measures primarily IQ and/or reading ability; and only from inspection that the induction section of the Cornell test Level Z measures mostly reading comprehension. Whimbey (1985) argued that various standardized achievement tests can serve as critical thinking tests, based on correlations between the two types of tests. For example, he reports correlations of .40 to .68 between the New Jersey College Basic Skills Placement Test and the Cornell tests. However, he gives no source for the data and no account of the 55% to 84% of nonoverlapping variance between the tests. This study provides evidence for evaluating criticisms like McPeck's and Whimbey's.

The inability properly to settle these issues given current evidence impinges adversely on educational practice. The critical thinking movement in many countries is motivated in part by a recognition that children learn large amounts of information, but learn less well how to evaluate and use that information. However, unless we understand the techniques for critical thinking assessment and the ways in which critical thinking tests work, we will not be able to determine whether the effort put into critical thinking instruction is leading to desired results. In addition, unless there is evidence on the validity of different critical thinking test formats, then policy decisions about what formats to use will not be grounded adequately.

Method

Definitions of Formats

Multiple choice versus constructed response. Traub's (1993) definitions were used to make the multiple-choice versus constructed-response format distinction: a multiple-choice test contains items "in which the examinee is required to choose an answer from a relatively small set of response options" (p. 29); a

constructed-response test "requires the examinee to compose an answer ... [ranging in length from] a word or a phrase, a number or a formula ... [to] a paragraph, an extended essay, a multistep solution to a mathematical or scientific problem" (pp. 29-30). All the multiple-choice measures used in this study were scored on the basis of a single correct answer for each item; the constructed-response measures contained items requiring from single-word responses to paragraph-length responses.

Among the 15 measures, were five pairs of what Traub and MacRury (1990) called "stem-equivalent" tests. A pair of multiple-choice and constructed-response tests is stem-equivalent if the tests are equivalent item-stem for item-stem, or "when, item for item, the tasks posed in each test are identical except for the response required" (Traub, 1993, p. 32, fn. #2). The pairs of measures were not, however, "scoring-equivalent" (Traub & MacRury, 1990). In multiple-choice testing, items usually are scored dichotomously as either correct or incorrect. All multiple-choice measures used in this study were scored this way, although three measures allowed for "scaled-answer" (Phillips & Patterson, 1987) scoring in which varying credit is assigned depending on the answer selected. In constructed-response testing, partial credit is often given for responses that are considered less than ideal, but somewhat meritorious. Among the five pairs of stem-equivalent measures, three of the constructed-response measures were scored so as to allow for zero, partial, and full credit. The other two constructed-response measures were scored dichotomously. However, the dichotomously scored constructed-response measures were still not scoring-equivalent to their stem-equivalent multiple-choice counterparts, because in each format different judgments were made to determine correctness. In the multiple-choice format, scores reflected whether or not examinees answered questions correctly. In the constructed-response versions, in addition to answering the same questions posed on the multiple-choice versions, examinees were required to justify their answers. Scores reflected whether or not their justifications were sound.

It makes sense to use different approaches and criteria in order to score stem-equivalent multiple-choice and constructed-response tests, because the formats are intended to provide different information for the examiner. Whether that difference in intent translates into measurement differences such that the two formats measure different psychological constructs is an issue to be studied here. It would bias unduly the outcome of the study, however, to adjust the scoring so that they are more similar than they would be in actual use. Hence scoring equivalence was not imposed on stem-equivalent pairs.

Narrative context versus nonnarrative context. The narrative and nonnarrative measures were differentiated, because different text types make distinct demands on readers' knowledge and expectations about reading, and consequently on their understanding of text. Narration is the most familiar to school students of all text types. It informs readers of what is happening by giving an account of events or actions. Typically, narrative text includes characters, plot, theme, and style. The narrative-context critical thinking tests studied here contained characters and themes; the nonnarrative tests did not.

Credibility judgment required versus no credibility judgment required. Credibility-judgment measures request that examinees decide on the believability of a

statement or of a person. This task is different from judging truth in that a statement may be believable but not true, or true but not believable. In credibility-judgment measures examinees often are presented pairs of statements. Each statement in a pair might be true, but examinees are asked to decide which is more credible. Such decisions are central to critical thinking and hence to many measures of the construct.

Common alternatives versus specific alternatives. The common-alternative versus specific-alternative distinction applies only to the multiple-choice measures. For common-alternative measures the same set of alternatives is used for all items. Such tests are quite prevalent in critical thinking testing. For instance, examinees may be given the choice for each item to judge whether evidence supports a given conclusion, goes against the conclusion, or is irrelevant to the conclusion. By comparison, in the specific-alternative measures examinees must comprehend and take on a new critical thinking task for each item or subset of items.

Justification requested versus no justification requested. The difference between the justification and no-justification measures resides in the former making a specific request for examinees to provide justification for their responses. Having to make their reasons explicit suggests a different and higher cognitive load, although examinees presumably consider justifications while taking the measures that request no justification.

Instruments

Fifteen critical thinking measures were used. Among these measures the following three comprised pairs of stem-equivalent multiple-choice and constructed-response formats: (a) Test on Appraising Observations, Part A; (b) Cornell Critical Thinking Test, Level X, Section I; and (c) Test of Inference Ability in Reading Comprehension. Abbreviated names for all measures and their format characteristics are provided in the first and second columns of Table 3.

Test on Appraising Observations (3 measures). This test contains questions set in the contexts of story narratives and assesses ability to judge the credibility of reports of observations. The multiple-choice format (Norris & King, 1983) consists of 50 items in two parts. Part A has 28 items set in the context of a traffic accident; Part B has 22 items set in the context of a river exploration. Each part was used as a separate measure in this study. Each item provides two statements spoken by characters who witnessed or who were involved in the accident or who made observations along the river. Examinees are to choose which, if either, of the statements they have more reason to believe at the time the statements are made. There are thus three alternatives for each item: (a) the first statement is more believable, (b) the second statement is more believable, or (c) neither statement is more believable, they are equally believable. The constructed-response format (Norris, 1986) is based on Part A of the multiple-choice format. It contains 25 items, and for each item examinees are to judge which, if either, of two statements is more believable. In addition, they are asked to justify their decision. Scores for each item represent whether or not the justification was sound. Some of the relevant factors in judging the statements are: the observer's expertise, alertness, and conflict of interest; the observation conditions; and the source of the observation and the statement reporting it.

Cornell Critical Thinking Test, Level X (2 measures). This test (Ennis & Millman, 1985) has four subsections, with 71 items. Items are cast in the context of a story about a group of scientific explorers from Earth who have gone to the planet Nicoma to search for some lost explorers. Section I, on induction, is the only one used in this study. It contains 23 items dealing with a health officer's hypothesis that the previous explorers are all dead. Different explorers report observations, and in the multiple-choice version examinees are to decide whether each observation report: (a) supports the health officers's idea that everyone in the first group is dead, (b) goes against the health officer's idea, or (c) does not help to decide on the hypothesis. A constructed-response version of Section I was designed for this study by altering the directions to request examinees to tell why they chose their answers and by providing an answer sheet to accommodate their justifications. The score for each item indicates whether or not the justification for the answer-choice is sound.

Test of Inference Ability in Reading Comprehension (6 measures). This test is designed to examine inference ability in the context of reading. It is a 36-item test in both multiple-choice (Phillips & Patterson, 1987) and constructed-response (Phillips, 1989) formats. Each format consists of three subtests based on either narrative, expository, or descriptive text. The first subtest in each format, *UFOs,* is an exposition about unusual atmospheric phenomena; *Money* is a description of the everyday use of money, of how it works, and of its historical development; and *The Wrong Newspapers* is a narrative about a mix-up in newspaper delivery. Each subtest consists of four to five paragraphs with questions after each for a total of 12 questions per subtest. In the multiple-choice format there are four alternatives per item. The constructed-response format uses the same 36 items; students are to formulate their own responses and justify them. Justifications are scored on a scale of 0 to 3 for each item.

Ennis-Weir Critical Thinking Essay Test. The Ennis-Weir (Ennis & Weir, 1985) is intended to evaluate ability to appraise an argument and to formulate a written argument in response. The test begins by asking examinees to read a letter to the editor of a fictional newspaper. In the letter, a proposal is made to end overnight parking on city streets, and a variety of arguments are offered in support of the proposal. Examinees are asked to write a letter evaluating the arguments in each paragraph and in the letter as a whole. The scoring system for each paragraph response is as follows: –1 for judging an argument incorrectly and/or showing bad judgment in justifying; 0 if no response if made; +1 if the argument is judged correctly but it is not justified; +2 if justified semiadequately; and +3 if justified adequately.

Test of Inductive Reasoning Principles. This test (Norris & Ryan, 1992) has 22 items written in the context of reports from biologists, language experts, physicians, and sociologists who are exploring the fictional planet Ladelin. In each question some data and either a generalization from or an explanation of the data is given by two different explorers. Examinees must choose which explorer, if either, gives stronger support for his or her conclusion.

Watson-Glaser Critical Thinking Appraisal. The Watson-Glaser (Watson & Glaser, 1980) comprises five subtests, each consisting of 16 items and covering the following aspects of critical thinking: inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments. In subtest I, the

ability to judge inferences made from a statement of facts is tested. There are five judgments from which to choose: true, probably true, insufficient data, probably false, or false. Subtest II provides five statements followed by a number of proposed assumptions. Examinees are to judge whether a person making a given statement is also making the proposed assumptions. Each assumption is to be judged independently from the others. In subtest III, examinees are to judge whether conclusions follow necessarily from given statements. Subtest IV tests for interpretation abilities. It contains five short paragraphs each followed by several suggested conclusions. Examinees are to judge whether each of the proposed conclusions follows beyond a reasonable doubt from the information. Finally, in subtest V, examinees are to judge whether given arguments are strong or weak.

Norris-Ryan Argument Analysis Test. This instrument (Norris & Ryan, 1989) is a 40-item multiple-choice test in five parts testing the ability to identify and distinguish arguments, reasons, and conclusions. Each section consists of a number of short paragraphs followed by questions with four alternative answers.

Participants

Participants were selected from grades 10, 11, and 12 in a large senior high school in central St. John's, Newfoundland. Participation was dependent on written consent from both students and at least one of their parents or guardians. A letter informing students about the study and what would be required if they chose to volunteer was distributed by the school principal to four classes at each grade level. The number of volunteers was 181, but complete data were collected on 172 students, distributed by grade and sex as in Table 2. Students were offered an honorarium for their time and effort. This sample represented a broad range of student abilities, with average school grades ranging from failing to 98%.

Procedure

The testing schedule was as follows: time was reserved each day after school for two weeks; each afternoon one test was given, sometimes including subtests; students were tested as a group in the school's cafeteria; at the end of the two weeks, make-up days were scheduled for students who had missed any testing. Following the recommendations of Traub and Fisher (1977) and Heim and Watts (1967) for the paired stem-equivalent multiple-choice and con-

Table 2
Subjects by Grade Level and Sex

	Grade			
Sex	10	11	12	Totals
Male	28	22	25	75
Female	43	29	25	97
Totals	71	51	50	172

structed-response measures, the constructed-response version of each pair was administered first.

The multiple-choice measures were machine scored using the answer keys and scoring procedures described in the manuals for each measure, with the exception of the multiple-choice version of the Reading Inference measures. For the Reading Inference measures, items were scored initially using two procedures, and the results compared: first, on the basis of the scale described in the manual, which rates answer choices on a range of quality; and, second, as right or wrong on the basis of whether or not the best answer was selected. The (attenuated) correlation between scores graded by both methods was 0.96. The scores from the right-wrong scoring procedure were used for this study on the assumption that it yielded the same information as the more complicated scaled-answer procedure.

For the constructed-response measures, the scoring systems described in the manuals were used, except for the Cornell Level X. The constructed-response format for this measure was developed for this study, so a scoring guide also was developed based on the justifications offered in the test manual (Ennis, Millman, & Tomko, 1985) for the keyed answers to each multiple-choice item.

Analysis

Descriptive statistics and a correlation matrix were computed for the 15 measures. A principal components analysis was performed and eigenvalues were examined to help determine a suitable number of factors. The viability of the various hypothetical factor models to account for the correlations among the measures was examined using EzPATH, Version 1.0 (Steiger, 1989), an implementation of structural equation modeling used by SYSTAT. Using the correlation matrix as input, EzPATH was directed to compute factor loadings with no constraints on the sizes or signs of the loadings. For each model EzPATH also was directed to compute a unique factor loading for each measure, again with the sizes and signs of the loadings unconstrained. For each model, solutions for both orthogonal and oblique common factors were attempted. The method of estimation first produced least squares estimates for factor loadings and correlations, and then used these estimates as input to a maximum likelihood estimation. In two cases it proved impossible to estimate the models because of insufficient constraints that led to out-of-bounds estimates of some factor loadings (loadings exceeding 1.0).

Three indices were used to judge the goodness of fit between the observed correlation matrix used as input and that computed from the estimated model. The null hypothesis of perfect match between the matrices was tested using the chi-squared statistic. The probability levels reported are the probabilities of obtaining values of χ^2 equal to or greater than the obtained values on the hypothesis that the computed matrices perfectly fit the observed matrix.

In addition to the test of perfect fit supplied by the chi-squared, degree of goodness of fit was measured by two indices: the Steiger-Lind RMS Index (R*) and the Adjusted Population Gamma Index (Γ_2). Values of R* less than .10 indicate a reasonably good fit, values below .05 indicate an excellent fit, and values less than .01 indicate an outstanding fit. Values of Γ_2 greater than .90 indicate a good fit and values greater than .95 indicate an excellent fit. The 90%

Table 3
Format Characteristics, Numbers of Items, Means, Standard Deviations, Alpha Reliabilities, and Correlation Matrix for All Measures

Measure	*Format*	*No. of items*	*Mean*	*S.D.*	*Alpha*	*No. of measure*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*
Observation Appraisal (A)	CR[1]/J[3]/NARR[7]/CRED[9]	25	16.0	3.86	.71	1	1.00														
Cornell, Level X (I)	CR/J/NARR/CRED	23	15.4	3.79	.72	2	.335	1.00													
Reading Inference *UFOs*	CR/NJ[4]/NONNARR[8]/NOCRED[10]	12	19.1	4.16	.43	3	.280	.277	1.00												
Reading Inference *Money*	CR/NJ/NONNARR/NOCRED	12	17.7	5.20	.58	4	.383	.209	.536	1.00											
Reading Inference *News*	CR/NJ/NARR/NOCRED	12	20.8	5.32	.59	5	.337	.205	.390	.294	1.00										
Ennis-Weir	CR/J/NONNARR/CRED	9	5.15	6.30	.72	6	.250	.323	.250	.310	.305	1.00									
Inductive Reasoning	MC[2]/CA[5]/NARR/CRED	22	10.6	3.15	.52	7	.309	.228	.174	.192	.210	.144	1.00								
Reading Inference *UFOs*	MC/SA[6]/NONNARR/NOCRED	12	8.29	2.16	.63	8	.360	.302	.314	.302	.316	.308	.314	1.00							
Reading Inference *Money*	MC/SA/NONNARR/NOCRED	12	7.71	1.70	.37	9	.220	.205	.145	.268	.255	.164	.128	.280	1.00						
Reading Inference *News*	MC/SA/NARR/NOCRED	12	7.87	2.02	.47	10	.268	.282	.240	.300	.344	.268	.224	.428	.417	1.00					
Watson-Glaser	MC/SA/NONNARR/NOCRED	80	48.6	8.06	.75	11	.257	.292	.377	.477	.342	.276	.258	.486	.343	.307	1.00				
Observation Appraisal (A)	MC/CA/NARR/CRED	28	18.5	3.30	.54	12	.584	.338	.359	.393	.246	.185	.312	.350	.359	.395	.290	1.00			
Observation Appraisal (B)	MC/CA/NARR/CRED	22	11.8	3.01	.52	13	.342	.291	.236	.285	.055	.183	.355	.320	.254	.234	.332	.565	1.00		
Argument Analysis	MC/SA/NONNARR/NOCRED	40	20.4	5.55	.76	14	.410	.352	.437	.525	.352	.407	.398	.508	.393	.388	.549	.461	.409	1.00	
Cornell, Level X (I)	MC/CA/NARR/CRED	23	15.9	3.14	.67	15	.288	.567	.222	.225	.276	.319	.342	.315	.242	.288	.301	.295	.303	.386	1.00

1 CR = constructed response
2 MC = multiple choice
3 J = justification requested
4 NJ = justification not requested
5 CA = common alternatives for all items
6 SA = specific alternatives for each item or item subset
7 NARR = items set in narrative context
8 NONNARR = items set in nonnarrative context
9 CRED = credibility judgment required
10 NOCRED = credibility judgment not required

confidence intervals on R* and Γ_2 were also computed in order to judge degree of confidence in the estimates of goodness of fit.

Results

Descriptive Statistics

Table 3 contains the format characteristics, number of items, mean, standard deviation, and alpha reliability for each measure, and the correlations among the measures. The highest reliability estimates, 0.75 for the 80-item Watson-Glaser test and 0.76 for the 40-item Argument Analysis test, are not high by the standards of many psychological tests. However, reported reliabilities for critical thinking tests tend to be low (Ennis & Norris, 1990).

Number of Factors

Table 4 contains the eigenvalues associated with the principal components. In determining the number of factors to extract, a variety of tests were applied to the pattern of eigenvalues, following the advice of Kim and Mueller (1978) who suggest seeking a convergence of a number of tests. The number of eigenvalues greater than one suggests four factors. If accounting for 5% of the variance is taken as a minimum amount for substantive importance of any component, then six factors seem correct. Applying the scree test suggests six, possibly four, or as few as two factors, depending on how the graph is viewed.

Theoretical consideration of the influence of format on the scores of the critical thinking measures used in this study suggested models with from three to six factors (see Table 1). The empirical data support testing models with either of these numbers of factors. Thus the judgment was made to test all the models proposed in Table 1.

Three-Factor Models

Table 5 presents the factor loadings, and, in the case of correlated factors, the correlations between format factors for Models 3A, 3B, and 3C. The top half of the table presents loadings for the orthogonal solutions, and the bottom half for the oblique solutions.

Almost without exception, measures load heavily on the Common Critical Thinking factor for each model. In the orthogonal solution to Model 3A the only multiple-choice measures to load on the Multiple-choice factor were the two Observation tests and the Induction test; of the constructed-response measures only the Reading measures loaded on the Constructed-response factor. The oblique solution to Model 3A produced much more clearly defined

Table 4
Eigenvalues Associated with Principal Components

Component	Eigenvalue	Component	Eigenvalue	Component	Eigenvalue
1	5.51	6	0.85	11	0.46
2	1.26	7	0.73	12	0.42
3	1.13	8	0.65	13	0.36
4	1.03	9	0.61	14	0.31
5	0.89	10	0.54	15	0.29

factors, although the Ennis-Weir and the Watson-Glaser did not load on the Common Critical Thinking factor.

In Model 3B neither the orthogonal solution nor the oblique solution yielded clearly defined Narrative and Nonnarrative factors. In Model 3C, no clear Credibility Judgment factor emerged in either the orthogonal solution or oblique solution; however, a No Credibility Judgment factor was present in both solutions.

The largest change in loadings between the orthogonal and oblique solutions was for Model 3A, where the loadings for the Multiple-choice and Constructed-response factors tended to be substantially higher for the solution with correlated factors. The Multiple-choice and Constructed-response factors were highly correlated in the oblique solution to Model 3A.

In Models 3B and 3C there was less of a marked difference in loadings between the orthogonal and oblique solutions. This is consistent with the low correlations between the Narrative and Nonnarrative factors and between the Credibility Judgment and No Credibility Judgment factors in the oblique solution.

Table 9 displays the results of the goodness of fit tests. In each case, the value of the χ^2 leads to a rejection of the hypothesis of perfect fit between the observed and computed correlation matrices. Examining R* indicates values between .05 and .10 for each model, indicating reasonably good fit. In each test the upper bound of the 90% confidence interval on R* was less than .10 increasing confidence in the judgment of a reasonably good fit. The Γ_2 values all are greater than .90, indicating good fit. However, in only one case is the lower bound of the 90% confidence interval greater than .90, reducing the confidence in the judgment of good fit for five of the models. Both the R* and Γ_2 indices indicate that the orthogonal solution to Model 3C fits the data best in this group. However, none of the models fits the data excellently.

Four-Factor Models

Table 6 presents the factor loadings, and in the case of correlated factors the correlations between format factors for Models 4A, 4B, and 4C.

In the orthogonal solution to Model 4A, measures tend to load heavily on their respective factors. In the Narrative and Nonnarrative factors, however, three measures do not load. An oblique solution to Model 4A with all correlations among factors unconstrained could not be tested without finding out-of-bounds estimates for some factor loadings. Attempts to solve this problem by placing further constraints on correlations among factors were made until the solution reported was found. The oblique solution is less easily interpreted than the orthogonal solution because of the negative loadings in the Narrative and Nonnarrative factors. The Credibility Judgment and No Credibility Judgment factors contain loadings very similar to the orthogonal solution.

In both Models 4B and 4C both the orthogonal and the oblique solutions yielded four clearly defined factors. In the oblique solutions the factors were highly intercorrelated, and there was no appreciable change in factor loadings over the orthogonal solutions.

Table 9 displays the results of the goodness of fit tests. In each case, the value of the χ^2 leads to a rejection of the hypothesis of perfect fit between the observed and computed correlation matrices. Examining R* and its confidence

Table 5
Factor Loadings for Three-Factor Models

Measure: Format	Model 3A Factors (1) Common Critical Thinking	(2) Multiple-Choice Format	(3) Constructed-Response Format	Model 3B Factors (1) Common Critical Thinking	(2) Narrative Context	(3) Non-Narrative Context	Model 3C Factors (1) Common Critical Thinking	(2) Credibility Judgment	(3) No Credibility Judgment
Orthogonal Solutions									
Observation A: CR/J/NARR/CRED	.58	*	−.02	.53	.34	*	.62	.21	*
Cornell: CR/J/NARR/CRED	.52	*	.02	.54	.04	*	.62	−.30	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	.51	*	.86	.49	*	.37	.42	*	.42
Reading Money: CR/NJ/NONNARR/NOCRED	.59	*	.27	.51	*	.76	.46	*	.49
Reading News: CR/NJ/NARR/NOCRED	.50	*	.16	.51	−.05	*	.38	*	.33
Ennis-Weir: CR/J/NONNARR/CRED	.48	*	.01	.48	*	.08	.43	−.24	*
Induction: MC/CA/NARR/CRED	.44	.17	*	.47	.06	*	.47	−.04	*
Reading UFOs: MC/SA/NONNARR/NOCRED	.64	.03	*	.66	*	−.04	.51	*	.35
Reading Money: MC/SA/NONNARR/NOCRED	.48	.04	*	.48	*	.02	.41	*	.23
Reading News: MC/SA/NARR/NOCRED	.56	−.03	*	.56	.10	*	.49	*	.22
Watson-Glaser: MC/SA/NONNARR/NOCRED	.64	.05	*	.63	*	.21	.45	*	.56
Observation A: MC/CA/NARR/CRED	.61	.32	*	.57	.82	*	.76	.50	*
Observation B: MC/CA/NARR/CRED	.46	.89	*	.48	.35	*	.59	.21	*
Argument: MC/SA/NONNARR/NOCRED	.78	.06	*	.76	*	.18	.64	*	.45
Cornell: MC/CA/NARR/CRED	.53	.07	*	.57	−.03	*	.64	−.40	*

Table 5 (continued)

Measure: Format		Model 3A Factors (1) Common Critical Thinking	(2) Multiple-Choice Format	(3) Constructed-Response Format	Model 3B Factors (1) Common Critical Thinking	(2) Narrative Context	(3) Non-Narrative Context	Model 3C Factors (1) Common Critical Thinking	(2) Credibility Judgment	(3) No Credibility Judgment
Oblique Solutions										
Observation A: CR/J/NARR/CRED		.49	*	.43	.51	.38	*	.52	.37	*
Cornell: CR/J/NARR/CRED		.22	*	.45	.55	.07	*	.61	.00	*
Reading UFOs: CR/NJ/NONNARR/NOCRED		.22	*	.57	.44	*	.44	.38	*	.54
Reading Money: CR/NJ/NONNARR/NOCRED		.24	*	.61	.45	*	.75	.41	*	.65
Reading News: CR/NJ/NARR/NOCRED		.11	*	.53	.50	−.01	*	.43	*	.25
Ennis-Weir: CR/J/NONNARR/CRED		.04	*	.53	.48	*	.10	.51	−.11	*
Induction: MC/CA/NARR/CRED		.19	.41	*	.47	.08	*	.48	.07	*
Reading UFOs: MC/SA/NONNARR/NOCRED		.16	.63	*	.66	*	.03	.60	*	.19
Reading Money: MC/SA/NONNARR/NOCRED		.24	.42	*	.46	*	.08	.43	*	.17
Reading News: MC/SA/NARR/NOCRED		.26	.48	*	.55	.13	*	.52	*	.16
Watson-Glaser: MC/SA/NONNARR/NOCRED		.08	.69	*	.61	*	.27	.53	*	.40
Observation A: MC/CA/NARR/CRED		.95	.31	*	.51	.86	*	.54	.82	*
Observation B: MC/CA/NARR/CRED		.47	.37	*	.46	.38	*	.49	.37	*
Argument: MC/SA/NONNARR/NOCRED		.23	.76	*	.74	*	.27	.68	*	.38
Cornell: MC/CA/NARR/CRED		.15	.49	*	.59	−.01	*	.66	−.08	*
Correlations Among Factors	(1)	—			—			—		
	(2)	*	—		*	—		*	—	
	(3)	*	.88	—	*	.25	—	*	.29	—

*Parameter constrained to be zero.

Table 6
Factor Loadings for Four-Factor Models

Measure: Format	Model 4A Factors (1) Narrative Context	(2) Non-Narrative Context	(3) Credibility Judgment	(4) No Credibility Judgment	Model 4B Factors (1) Multiple-Choice Common Alternative	(2) Multiple-Choice Specific Alternative	(3) Constructed-Response Justification	(4) Constructed-Response No Justification	Model 4C Factors (1) Multiple-Choice Narrative	(2) Multiple-Choice Non-Narrative	(3) Constructed-Response Narrative	(4) Constructed Response Non-Narrative
Orthogonal Solutions												
Observation A: CR/J/NARR/CRED	.51	*	.33	*	*	*	.51	*	*	*	.74	*
Cornell: CR/J/NARR/CRED	.17	*	.69	*	*	*	.66	*	*	*	.45	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	*	.47	*	.41	*	*	*	.84	*	*	*	.66
Reading Money: CR/NJ/NONNARR/NOCRED	*	.72	*	.42	*	*	*	.64	*	*	*	.81
Reading News: CR/NJ/NARR/NOCRED	.02	*	*	.50	*	*	*	.46	*	*	.45	*
Ennis-Weir: CR/J/NONNARR/CRED	*	.21	.40	*	*	*	.49	*	*	*	*	.38
Induction: MC/CA/NARR/CRED	.22	*	.38	*	.48	*	*	*	.48	*	*	*
Reading UFOs: MC/SA/NONNARR/NOCRED	*	.02	*	.69	*	.68	*	*	*	.66	*	*
Reading Money: MC/SA/NONNARR/NOCRED	*	.03	*	.51	*	.52	*	*	*	.48	*	*
Reading News: MC/SA/NARR/NOCRED	.17	*	*	.52	*	.56	*	*	.46	*	*	*
Watson-Glaser: MC/SA/NONNARR/NOCRED	*	.25	*	.64	*	.69	*	*	*	.72	*	*
Observation A: MC/CA/NARR/CRED	.97	*	.24	*	.72	*	*	*	.76	*	*	*
Observation B: MC/CA/NARR/CRED	.50	*	.32	*	.76	*	*	*	.70	*	*	*
Argument: MC/SA/NONNARR/NOCRED	*	.30	*	.69	*	.76	*	*	*	.78	*	*
Cornell: MC/CA/NARR/CRED	.10	*	.78	*	.44	*	*	*	.46	*	*	*

Table 6 (continued)

Measure: Format	Model 4A Factors (1) Narrative Context	(2) Non-Narrative Context	(3) Credibility Judgment	(4) No Credibility Judgment	Model 4B Factors (1) Multiple-Choice Common Alternative	(2) Multiple-Choice Specific Alternative	(3) Constructed-Response Justification	(4) Constructed-Response No Justification	Model 4C Factors (1) Multiple-Choice Narrative	(2) Multiple-Choice Non-Narrative	(3) Constructed-Response Narrative	(4) Constructed Response Non-Narrative
Oblique Solutions												
Observation A: CR/J/NARR/CRED	.33	*	.57	*	*	*	.64	*	*	*	.64	*
Cornell: CR/J/NARR/CRED	−.18	*	.67	*	*	*	.55	*	*	*	.54	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	*	.35	*	.49	*	*	*	.68	*	*	*	.66
Reading Money: CR/NJ/NONNARR/NOCRED	*	.81	*	.50	*	*	*	.73	*	*	*	.74
Reading News: CR/NJ/NARR/NOCRED	−.03	*	*	.51	*	*	*	.53	*	*	.46	*
Ennis-Weir: CR/J/NONNARR/CRED	*	.11	.43	*	*	*	.44	*	*	*	*	.48
Induction: MC/CA/NARR/CRED	.03	*	.47	*	.50	*	*	*	.50	*	*	*
Reading UFOs: MC/SA/NONNARR/NOCRED	*	−.06	*	.68	*	.65	*	*	*	.64	*	*
Reading Money: MC/SA/NONNARR/NOCRED	*	.01	*	.50	*	.49	*	*	*	.48	*	*
Reading News: MC/SA/NARR/NOCRED	.10	*	*	.57	*	.55	*	*	.54	*	*	*
Watson-Glaser: MC/SA/NONNARR/NOCRED	*	.17	*	.65	*	.67	*	*	*	.68	*	*
Observation A: MC/CA/NARR/CRED	.62	*	.66	*	.72	*	*	*	.72	*	*	*
Observation B: MC/CA/NARR/CRED	.31	*	.55	*	.63	*	*	*	.59	*	*	*
Argument: MC/SA/NONNARR/NOCRED	*	.17	*	.76	*	.81	*	*	*	.82	*	*
Cornell: MC/CA/NARR/CRED	−.30	*	.72	*	.56	*	*	*	.56	*	*	*
Correlations Among Factors (1)	—				—				—			
(2)	.24	—			.81	—			.84	—		
(3)	*	*	—		.98	.83	—		.99	.82	—	
(4)	*	*	.79	—	.63	.82	.77	—	.68	.83	.81	—

* Parameter constrained to be zero.

interval indicates that the oblique models 4B and 4C have reasonably good fit with the data, and that Model 4A is bordering on an excellent fit. Examining Γ_2 values and its confidence interval leads to the same conclusion. However, the lower bounds of the 90% confidence interval is less than .90 for both models 4B and 4C, reducing the confidence in the judgment of good fit. As with the three-factor models, none of the four-factor models fits the data excellently, although the oblique solution to Model 4A is close.

Five-Factor Models

Table 7 presents the factor loadings, and in the case of correlated factors the correlations between format factors for Models 5A, 5B, and 5C.

In the orthogonal solution to Model 5A, all measures load heavily on the Common Critical Thinking factor. For the Narrative factor, the Reading News measure loads negatively in both its constructed-response and multiple-choice formats, although the loading for the multiple-choice measure is close to zero. In the Nonnarrative factor, all measures load positively, although the loading for the Reading Money multiple-choice measure is close to zero. Both the Credibility Judgment and No Credibility Judgment factors are bipolar.

The oblique solution to Model 5A could not be tested with all correlations among factors unconstrained without finding out-of-bounds estimates for some factor loadings. Attempts to solve this problem by placing further constraints on the model were made until the model reported was reached. In this solution, the bipolarity in the Credibility Judgment and No Credibility Judgment factors remains.

In the orthogonal solutions to both Models 5B and 5C, there were some near-zero loadings for most factors. In the oblique solutions, the factors were highly intercorrelated, especially for Model 5B. The oblique solution to Model 5B tended to have smaller loadings on the common Critical Thinking factor, and higher loadings on the other four factors, compared with the orthogonal solution to Model 5B. In both Models 5B and 5C, there were fewer near-zero loadings in the oblique solutions than in the orthogonal solutions.

Table 9 displays the results of the goodness of fit tests. In the cases of Models 5B and 5C, the values of the χ^2 lead to a rejection of the hypotheses of perfect fit between the observed and computed correlation matrices. Examining R* for these models indicates reasonably good fit with the data. Examining Γ_2 values leads to the same conclusion. However, the lower bounds of the 90% confidence interval for the orthogonal solutions are less than .90, reducing the confidence in the judgment of good fit.

In the case of Model 5A, the value of the χ^2 means that the hypothesis of perfect fit between the observed and computed correlation matrices cannot be rejected. Furthermore, the values and confidence intervals for the R* and Γ_2 indices support the conclusion of an excellent fit to the data for both the orthogonal and oblique solutions.

Six-Factor Models

Table 8 presents the factor loadings, and in the case of correlated factors the correlations between format factors for Models 6A, 6B, and 6C. Note that oblique solutions for Models 6A and 6C were not possible because the models provided insufficient constraints for EzPATH to reach convergence.

Table 7
Factor Loadings for Five-Factor Models

Measure: Format	Model 5A Factors					Model 5B Factors					Model 5C Factors				
	(1) Common Critical Thinking	(2) Narrative Context	(3) Non-Narrative Context	(4) Credibility Judgment	(5) No Credibility Judgment	(1) Common Critical Thinking	(2) Multiple-Choice Common Alternative	(3) Multiple-Choice Specific Alternative	(4) Constructed-Response Justification	(5) Constructed-Response No Justification	(1) Common Critical Thinking	(2) Multiple-Choice Narrative	(3) Multiple-Choice Non-Narrative	(4) Constructed-Response Narrative	(5) Constructed-Response Non-Narrative
Orthogonal Solutions															
Observation A: CR/J/NARR/CRED	.62	.14	*	−.17	*	.57	*	*	−.02	*	.60	*	*	.04	*
Cornell: CR/J/NARR/CRED	.54	.25	*	.39	*	.51	*	*	.09	*	.55	*	*	−.09	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	.48	*	.50	*	−.21	.51	*	*	*	.86	.52	*	*	*	.29
Reading Money: CR/NJ/NONNARR/NOCRED	.48	*	.58	*	−.00	.60	*	*	*	.27	.58	*	*	*	.82
Reading News: CR/NJ/NARR/NOCRED	.69	−.70	*	*	−.18	.49	*	*	*	.16	.51	*	*	.86	*
Ennis-Weir: CR/J/NONNARR/CRED	.42	*	.21	.23	*	.47	*	*	.88	*	.48	*	*	*	.04
Induction: MC/CA/NARR/CRED	.44	.12	*	.07	*	.45	.15	*	*	*	.45	.15	*	*	*
Reading UFOs: MC/SA/NONNARR/NOCRED	.54	*	.18	*	.33	.63	*	.10	*	*	.60	*	.25	*	*
Reading Money: MC/SA/NONNARR/NOCRED	.45	*	.03	*	.34	.47	*	.19	*	*	.46	*	.14	*	*
Reading News: MC/SA/NARR/NOCRED	.54	−.04	*	*	.29	.55	*	.84	*	*	.57	−.05	*	*	*
Watson-Glaser: MC/SA/NONNARR/NOCRED	.51	*	.40	*	.27	.65	*	−.06	*	*	.57	*	.56	*	*
Observation A: MC/CA/NARR/CRED	.73	.37	*	−.38	*	.61	.31	*	*	*	.64	.29	*	*	*
Observation B: MC/CA/NARR/CRED	.51	.41	*	−.12	*	.49	.87	*	*	*	.49	.87	*	*	*
Argument: MC/SA/NONNARR/NOCRED	.64	*	.39	*	.27	.80	*	−.06	*	*	.74	*	.23	*	*
Cornell: MC/CA/NARR/CRED	.58	.18	*	.51	*	.53	.05	*	*	*	.56	.04	*	*	*

Table 7 (continued)

Measure: Format	Model 5A Factors (1) Common Critical Thinking	(2) Narrative Context	(3) Non-Narrative Context	(4) Credibility Judgment	(5) No Credibility Judgment	Model 5B Factors (1) Common Critical Thinking	(2) Multiple-Choice Common Alternative	(3) Multiple-Choice Specific Alternative	(4) Constructed-Response Justification	(5) Constructed-Response No Justification	Model 5C Factors (1) Common Critical Thinking	(2) Multiple-Choice Narrative	(3) Multiple-Choice Non-Narrative	(4) Constructed-Response Narrative	(5) Constructed-Response Non-Narrative
Oblique Solutions															
Observation A: CR/J/NARR/CRED	.63	−.18	*	−.09	*	−.39	*	*	.51	*	.46	*	*	.89	*
Cornell: CR/J/NARR/CRED	.50	−.21	*	.46	*	−.00	*	*	.66	*	.62	*	*	.06	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	.49	*	.46	*	−.15	−.22	*	*	*	.65	.41	*	*	*	.46
Reading Money: CR/NJ/NONNARR/NOCRED	.48	*	.63	*	−.03	−.24	*	*	*	.68	.37	*	*	*	.84
Reading News: CR/NJ/NARR/NOCRED	.72	.68	*	*	−.13	−.08	*	*	*	.53	.47	*	*	.14	*
Ennis-Weir: CR/J/NONNARR/CRED	.40	*	.21	.26	*	.07	*	*	.51	*	.49	*	*	*	.15
Induction: MC/CA/NARR/CRED	.43	−.13	*	.14	*	−.08	.48	*	*	*	.47	.11	*	*	*
Reading UFOs: MC/SA/NONNARR/NOCRED	.53	*	.16	*	.40	−.11	*	.64	*	*	.59	*	.25	*	*
Reading Money: MC/SA/NONNARR/NOCRED	.44	*	.05	*	.29	−.22	*	.45	*	*	.40	*	.26	*	*
Reading News: MC/SA/NARR/NOCRED	.53	.01	*	*	.27	−.22	*	.51	*	*	.52	.17	*	*	*
Watson-Glaser: MC/SA/NONNARR/NOCRED	.49	*	.39	*	.32	−.05	*	.69	*	*	.51	*	.48	*	*
Observation A: MC/CA/NARR/CRED	.75	−.44	*	−.28	*	−.86	.52	*	*	*	.47	.87	*	*	*
Observation B: MC/CA/NARR/CRED	.50	−.43	*	−.01	*	−.36	.50	*	*	*	.44	.41	*	*	*
Argument: MC/SA/NONNARR/NOCRED	.62	*	.38	*	.30	−.18	*	.78	*	*	.65	*	.49	*	*
Cornell: MC/CA/NARR/CRED	.53	−.13	*	.57	*	.10	.74	*	*	*	.67	−.02	*	*	*
Correlations Among Factors															
(1)	—					—					—				
(2)	*	—				*	—				*	—			
(3)	*	−.06	—			*	.75	—			*	.30	—		
(4)	*	*	*	—		*	.96	.79	—		*	.47	.19	—	
(5)	*	*	*	.26	—	*	.54	.81	.71	—	*	.30	.66	.28	—

* Parameter constrained to be zero.

Table 8
Factor Loadings for Six-Factor Models

Measure: Format	Model 6A Factors (1) Mul. Ch.	(2) Const. Resp.	(3) Narr.	(4) Non-Narr.	(5) Cred. Judg.	(6) No Cred. Judg.	Model 6B Factors (1) Narr.	(2) Non-Narr.	(3) Mul. Ch. Comm. Alt.	(4) Mul. Ch. Spec. Alt.	(5) Const. Resp. Just.	(6) Const. Resp. No Just.	Model 6C Factors (1) Cred. Judg.	(2) No Cred. Judg:	(3) Mul. Ch. Narr.	(4) Mul. Ch. Non-Narr.	(5) Const. Resp. Narr.	(6) Const. Resp. Non-Narr.
Orthogonal Solutions																		
Observation A: CR/J/NARR/CRED	*	−.03	.58	*	.22	*	.62	*	*	*	.03	*	.62	*	*	*	.78	*
Cornell: CR/J/NARR/CRED	*	.07	.30	*	.68	*	.56	*	*	*	.16	*	.65	*	*	*	−.09	*
Reading UFOs: CR/NJ/NONNARR/NOCRED	*	.82	*	.48	*	.25	*	.50	*	*	*	.71	*	.52	*	*	*	.28
Reading Money: CR/NJ/NONNARR/NOCRED	*	.14	*	.64	*	.30	*	.62	*	*	*	.29	*	.60	*	*	*	.78
Reading News: CR/NJ/NARR/NOCRED	*	.20	.08	*	*	.55	.34	*	*	*	*	.25	*	.49	*	*	.14	*
Ennis-Weir: CR/J/NONNARR/CRED	*	−.03	*	.36	.34	*	*	.45	*	*	.78	*	.38	*	*	*	*	.07
Induction: MC/CA/NARR/CRED	.46	*	.14	*	.19	*	.47	*	−.10	*	*	*	.46	*	−.01	*	*	*
Reading UFOs: MC/SA/NONNARR/NOCRED	.37	*	*	.21	*	.45	*	.52	*	.33	*	*	*	.62	*	.37	*	*
Reading Money: MC/SA/NONNARR/NOCRED	.25	*	*	.09	*	.44	*	.33	*	.41	*	*	*	.52	*	−.10	*	*
Reading News: MC/SA/NARR/NOCRED	.19	*	.21	*	*	.52	.35	*	*	.62	*	*	*	.53	.12	*	*	*
Watson-Glaser: MC/SA/NONNARR/NOCRED	.30	*	*	.39	*	.42	*	.66	*	.19	*	*	*	.69	*	.16	*	*
Observation A: MC/CA/NARR/CRED	.26	*	.94	*	-.03	*	.78	*	.62	*	*	*	.68	*	.73	*	*	*
Observation B: MC/CA/NARR/CRED	.48	*	.40	*	.08	*	.56	*	.21	*	*	*	.54	*	.27	*	*	*
Argument: MC/SA/NONNARR/NOCRED	.44	*	*	.50	*	.37	*	.79	*	.19	*	*	*	.77	*	.08	*	*
Cornell: MC/CA/NARR/CRED	.24	*	.19	*	.66	*	.67	*	−.38	*	*	*	.70	*	−.25	*	*	*

Table 8 (continued)

Measure: Format	Model 6A Factors (1) Mul. Ch.	(2) Const. Resp.	(3) Narr.	(4) Non-Narr.	(5) Cred. Judg.	(6) No Cred. Judg.	Model 6B Factors (1) Narr.	(2) Non-Narr.	(3) Mul. Ch. Comm. Alt.	(4) Mul. Ch. Spec. Alt.	(5) Const. Resp. Just.	(6) Const. Resp. No Just.	Model 6C Factors (1) Cred. Judg.	(2) No Cred. Judg.	(3) Mul. Ch. Narr.	(4) Mul. Ch. Non-Narr.	(5) Const. Resp. Narr.	(6) Const. Resp. Non-Narr.
Oblique Solutions																		
Observation A: CR/J/NARR/CRED							.39	*	*	*	.56	*						
Cornell: CR/J/NARR/CRED							−.11	*	*	*	.68	*						
Reading UFOs: CR/NJ/NONNARR/NOCRED							*	.38	*	*	*	.55						
Reading Money: CR/NJ/NONNARR/NOCRED							*	.68	*	*	*	.49						
Reading News: CR/NJ/NARR/NOCRED							.01	*	*	*	*	.63						
Ennis-Weir: CR/J/NONNARR/CRED							*	.13	*	*	.46	*						
Induction: MC/CA/NARR/CR	UNABLE TO TEST MODEL						.06	*	.47	*	*	*	UNABLE TO TEST MODEL					
Reading UFOs: MC/SA/NONNARR/NOCRED							*	.08	*	.67	*	*						
Reading Money: MC/SA/NONNARR/NOCRED							*	.10	*	.49	*	*						
Reading News: MC/SA/NARR/NOCRED							.14	*	*	.58	*	*						
Watson-Glaser: MC/SA/NONNARR/NOCRED							*	.31	*	.60	*	*						
Observation A: MC/CA/NARR/CRED							.67	*	.61	*	*	*						
Observation B: MC/CA/NARR/CRED							.35	*	.53	*	*	*						
Argument: MC/SA/NONNARR/NOCRED							*	.35	*	.71	*	*						
Cornell: MC/CA/NARR/CRED							−.28	*	.78	*	*	*						
Correlations Among Factors (1)							—											
(2)							.37	—										
(3)							*	*	—									
(4)							*	*	.75	—								
(5)							*	*	.92	.78	—							
(6)							*	*	.62	.78	.76	—						

Table 9
Fit Indices for Hypothesized Factor Models

Model	*df*	*χ2 (prob.)*	*Steiger-Lind Adjusted RMS Index (R*) (90% confidence interval)*	*Adjusted Population Gamma Index (Γ_2) (90% confidence interval)*
Three-Factor Models				
Orthogonal factors				
3A	75	164.3 (.00)	.080 (.062-.097)	.904 (.862-.941)
3B	75	134.1 (.00)	.065 (.045-.084)	.935 (.896-.968)
3C	75	127.3 (.00)	.062 (.042-.081)	.941 (.902-.973)
Oblique Factors				
3A	74	143.7 (.00)	.073 (.054-.091)	.920 (.879-.955)
3B	74	129.0 (.00)	.063 (.043-.082)	.939 (.899-.972)
3C	74	127.4 (.00)	.064 (.043-.082)	.938 (.898-.971)
Four-Factor Models				
Orthogonal factors				
4A	75	198.3 (.00)	.087 (.069-.104)	.889 (.845-.927)
4B	90	425.9 (.00)	.154 (.140-.168)	.704 (.662-.746)
4C	90	454.9 (.00)	.161 (.147-.176)	.682 (.640-.724)
Oblique Factors				
4A	73	109.5 (.00)	.050 (.024-.070)	.961 (.925-.991)
4B	84	171.2 (.00)	.076 (.058-.093)	.914 (.875-.948)
4C	84	180.5 (.00)	.079 (.062-.096)	.906 (.867-.941)
Five-Factor Models				
Orthogonal Factors				
5A	60	73.1 (.12)	.034 (.000-.060)	.982 (.944-1.000)
5B	75	151.3 (.00)	.072 (.053-.090)	.921 (.880-.956)
5C	75	159.1 (.00)	.077 (.059-.095)	.910 (.868-.946)
Oblique Factors				
5A	58	71.7 (.11)	.034 (.000-.060)	.982 (.943-1.000)
5B	69	109.3 (.00)	.055 (.031-.075)	.953 (.914-.985)
5C	69	117.0 (.00)	.061 (.039-.081)	.942 (.902-.976)
Six-Factor Models				
Orthogonal Factors				
6A	60	152.7 (.00)	.092 (.072-.110)	.874 (.823-.920)
6B	75	225.6 (.00)	.096 (.079-.113)	.865 (.820-.906)
6C	75	235.0 (.00)	.099 (.083-.116)	.856 (.811-.898)
Oblique Factors				
6A			UNABLE TO TEST MODEL	
6B	68	99.0 (.01)	.046 (.015-.068)	.968 (.930-.997)
6C			UNABLE TO TEST MODEL	

In the orthogonal solution to Model 6A, the Multiple-choice and No Credibility Judgment factors were the most clearly defined. The other four factors each contained a mixture of substantial and almost zero loadings. In the orthogonal solution to Model 6B all factors are reasonably clearly defined; the Multiple-choice Common Alternative factor was bipolar. The oblique solution to Model 6B removed the bipolarity on the Multiple-choice Common Alternative factor and introduced it into the Narrative Context factor. Substantial correlations were found among those factors that were allowed to be correlated in the analysis. In the orthogonal solution to Model 6C, the Credibility Judgment and No Credibility Judgment factors were the most clearly defined. Each of the remaining four factors contained at least one near-zero loading, and the Multiple-choice Narrative factor was bipolar.

Table 9 displays the results of the goodness of fit tests. For all the models tested, the value of the χ^2 leads to a rejection of the hypothesis of perfect fit between the observed and computed correlation matrices. Only in the case of the oblique solution to Model 6B did the R* and Γ_2 values indicate excellent fit, but this judgment must be tempered by the fact that the lower bound of the 90% confidence interval for Γ_2 is less than .95, and the upper bound of the interval on R* is greater than .05.

Discussion

Twelve models were hypothesized to explain the correlations among 15 measures of critical thinking on the basis of the different formats of the measures. Each of the models was translated into a hypothesized factor structure that was then tested using confirmatory factor analysis. Except for two cases (Models 6A and 6C), each factor structure was tested, first, with the factors constrained to be orthogonal and, second, with at least some of the factors allowed to be correlated. Oblique solutions to Models 6A and 6C could not be found.

Only one model, 5A, was able to pass the χ^2 test of perfect fit between the observed correlations among the measures and the correlations implied by the hypothesized factor structure. The model also showed excellent fit to the data according to both the Steiger-Lind Adjusted RMS Index and the Adjusted Population Gamma Index. According to both these indices it was the best fitting model. The orthogonal and oblique solutions performed the same according to all three indices. The factor loadings also were very similar for both solutions, so the criterion of theoretical simplicity would suggest support of the orthogonal solution.

Model 5A postulated five factors: a Common Critical Thinking factor, and four simple format factors: Narrative Context, Nonnarrative Context, Credibility Judgment Required, and No Credibility Judgment Required. The Common Critical Thinking factor received substantial loadings from all measures, which suggests that all these measures to a large extent test for the same construct. This is an interesting and perhaps happy finding given the diversity of the measures and the conceptual models on which they are based and the fact that they purport to measure aspects of the same construct, critical thinking. Of the remaining four factors two refer to the context (narrative versus nonnarrative) in which the items were set, and two to the nature of the task presented to examinees (credibility judgment required versus no credibility judgment re-

quired). Three of these four factors were bipolar. This presents interpretation problems because conceptually each of the factors was designed as a contrasting pair. Thus, for example, what does it mean to have a negative loading on the Narrative Context factor as opposed to a positive loading on the Nonnarrative Context factor? The bipolarity challenges the concept labels that were placed on the factors and used to construct the Model 5A hypothesis. Thus although the result was a model that fitted the data well, the underlying empirical structure of the model perhaps is not explained well by the conceptual structure that was used to design it.

Among the three-factor and four-factor models, Models 3C (orthogonal solution), and 4A (oblique solution) showed the best fit to the data on the basis of both the Steiger-Lind and Population Gamma indices. We see in Models 3C and 4A only factors also contained in Model 5A. Taken together these results suggest the rejection of the Multiple-choice and Constructed-response factors as viable considerations in attempting to explain the correlations among the set of critical thinking measures studied here. This conclusion is significant, because most of the debate over the format for critical thinking tests has centered around the relative adequacy of multiple-choice and constructed-response approaches. The results of this study suggest that much of this debate is groundless because that particular format difference had little distinguishable effect on performance in the measures studied.

Models 4B, 5B, and 6B (oblique solutions) were exceptions to this finding. These models did not fit the data as well as Model 5A, but showed a reasonably good fit on the basis of the Steiger-Lind and Population Gamma indices. The models employed compound Multiple-choice and Constructed-response factors. Thus an effect on performance for the multiple-choice versus constructed-response format difference appears when other task differences in the multiple-choice and constructed-response formats are represented in the factor structure. Given this finding, and the finding that Model 3A, which contained simple multiple-choice and constructed-response factors, tested less well than Models 3B and 3C, it is reasonable to conclude that the overriding determiners in Models 4B, 5B, and 6B were other than the multiple-choice and constructed-response formats.

One of the major limitations of this study arises from the selection of critical thinking measures examined. The 15 measures originate from only four organizations. The precaution was taken to test the hypothesis that four organization-of-origin factors could account for the data. The orthogonal solution to this model had to be rejected on the basis of all three indices of fit. The oblique solution also had to be rejected, although it showed a reasonably good fit to the data on the basis of the Steiger-Lind index and fitted the data as well as some of the other rejected models. Nevertheless, the measures do represent a narrowness of origin that limits the generalizability of the results. The narrowness was hardly avoidable, however, because few critical thinking tests are available commercially (Norris & Ennis, 1989), and those selected for the study represent the available measures well.

In addition to limited range of origin, the instruments also represent a limited range of content. In particular, there were no subject-specific critical thinking measures studied, that is, measures of critical thinking in, for ex-

ample, science, mathematics, literature, or history. It is impossible to predict how the results would have turned out if such measures had been available and represented in the study. Subject matter content would have suggested other explanatory factors to explore. Little is known about critical thinking testing in specific subjects.

The study was limited to performance by senior high school students, but the range of advertised applicability for some of the measures extended below or above high school grades, or both. Thus the results cannot be generalized automatically to all subjects with whom these measures might be used. Perhaps, for instance, younger students are differently influenced by format effects of the type studied here than older students.

Some precaution is warranted in interpreting the results of confirmatory factor analysis in terms of measurement theory concepts. There is a similarity between the classical test theory decomposition of variable scores into true and error scores, and the factor analytic decomposition into common and unique factors. It is this similarity that has led to the use of confirmatory factor analysis as a technique for checking hypotheses about test equivalence. Thus the definition of parallel measures can be translated into factor analytic language as: "two tests are said by psychometricians to be parallel if they share equal amounts of a common factor, and each also has the same amount of specific variance" (Loehlin, 1987, p. 85). By extension, tests would be tau-equivalent if they share equal amounts of a common factor, but do not have the same amount of specific variance; and tests would be congeneric if they share a common factor, but not to the same degree, and do not have the same amount of specific variance. On this interpretation this study conducted tests of whether groups of measures were congeneric.

There are some theoretical problems with equating classical test theory and common factor theory. Even though the foundations of common factor theory and classical test theory have been laid for almost 90 years, there is no full understanding of the theoretical and empirical relationships between partitioning scores into true and error scores on the one hand, and common and unique factors on the other (McDonald, 1985). The problem arises because it is not theoretically sound to assume that all the unique factor variance in scores is attributable to error. For example, part of the unique factor variance of a test could be attributable to factors that the test shares with other tests that currently are not under consideration. This means that when confirmatory factor analysis is used to check hypotheses of test equivalence, it must be understood that the analysis does not distinguish between variance that is unique to a test in the context of the current analysis but that it shares with other tests not under consideration, and variance that is error and thus uncorrelated with anything else.

Having examined the above limitations, it also must be said that the factor structure of Model 5A makes conceptual and theoretical sense. The existence of a Common Critical Thinking factor should have been expected if the measures were to have theoretical unity. All 15 measures purport to examine some aspects of critical thinking, and it is reasonable to assume that different aspects of critical thinking are highly related to one another. A key word in the previous sentence is *purport*. It must be kept in mind that there was no inde-

pendent test of the construct validity of the measures used in this study. So the finding of a common factor, strictly speaking, demonstrates only that all the measures test *something* in common. The finding does not, in addition, mean that this something is critical thinking. Calling it *critical thinking* is, however, a reasonable hypothesis given two facts: each measure comes with evidence from its developer that it measures critical thinking, and the measures demonstrably test something in common.

The Narrative and Nonnarrative Context factors also make sense. It is a consistent finding that narrative texts are easier to understand than nonnarratives for both adults and children (Bereiter & Scardamalia, 1982; Bock & Brewer, 1985). Narrative is the text type used for initial reading instruction and remains thereafter the type most comprehensible to readers. Clearly initial comprehension of the text with clarity and precision is fundamental to thinking critically about the content.

The Credibility Judgment and No Credibility Judgment factors also make sense as determiners of performance on these measures. Credibility judgment is a sophisticated form of thought requiring knowledge of specific criteria of adjudication and skill in the use and weighing and balancing of these criteria. If an examinee does not know or think to use suitable criteria, then he or she cannot perform well.

This study presents solid evidence to help temper much of the rhetoric surrounding critical thinking testing (McPeck, 1981; Whimbey, 1985). In particular the results are relevant to the dispute over the relative merits of multiple-choice and constructed-response critical thinking measures. In line with Hogan's (1981) conclusion from an examination of this format difference in other than critical thinking tests, the two formats appear to be equivalent or nearly so. In line with Traub and MacRury (1990), any difference that exists is slight. On the other hand, the type of text in which items are cast and whether or not examinees are asked to judge credibility surface as important factors. The latter task is highly significant because facility with it is at the foundation of all critical thinking.

Acknowledgment

Preparation of this manuscript was supported by a grant from the Social Sciences and Humanities Research Council of Canada, grant number 410-89-0052.

References

Arter, J., & Salmon, J. (1987). *Assessing higher order thinking skills: A consumer's guide*. Portland, OR: Northwest Regional Educational Laboratory.

Barnett-Foster, D., & Nagy, P. (1995). A comparison of undergraduate test response strategies for multiple-choice and constructed-response questions. *Alberta Journal of Educational Research, 41*, 18-35.

Bennett, R.E., & Ward, W.C. (Eds.). (1993). *Construction versus choice in cognitive measurement*. Hillsdale, NJ: Erlbaum.

Bereiter, C., & Scardamalia, M. (1982). From conversation to composition. The role of instruction in a developmental process. In R. Glaser (Ed.), *Advances in instructional psychology* (vol. 2, pp. 1-64). Hillsdale, NJ: Erlbaum.

Black, M. (1946). *Critical thinking*. Englewood Cliffs, NJ: Prentice-Hall.

Bock, J.K., & Brewer, W.F. (1985). Discourse structure and mental models. In T.H. Carr (Ed.), *The development of reading skills* (pp. 55-75). San Francisco, CA: Jossey-Bass.

Dewey, J. (1933). *How we think*. Boston: Heath.

Ennis, R.H. (1962). A concept of critical thinking. *Harvard Educational Review, 32*, 81-111.

Ennis, R.H. (1981). Rational thinking and educational practice. In J.F. Soltis (Ed.), *Philosophy of education* (pp. 143-83). Chicago, IL: National Society for the Study of Education.
Ennis, R.H. (1984). Problems in testing informal logic/critical thinking/reasoning ability. *Informal Logic, 6*(1), 3-9.
Ennis, R.H., & Millman, J. (1985). *Cornell critical thinking test level X*. Pacific Grove, CA: Midwest.
Ennis, R.H., Millman, J., & Tomko, T.N. (1985). *Cornell critical thinking tests levels X and Z manual*. Pacific Grove, CA: Midwest.
Ennis, R.H., & Norris, S.P. (1990). Critical thinking assessment: Status, issues, needs. In S. Legg & J. Algina (Eds.), *Cognitive assessment of language and math outcomes* (pp. 1-42). Norwood, NJ: Ablex.
Ennis, R.H., & Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific Grove, CA: Midwest.
Heim, A.W., & Watts, K.P. (1967). An experiment on multiple-choice versus open-ended answering in a vocabulary test. *British Journal of Educational Psychology, 37*, 339-346.
Hogan, T.P. (1981). *Relationship between free-response and choice-type tests of achievement: A review of the literature*. Green Bay, WI: University of Wisconsin. (ERIC Document Reproduction Service No. ED 224 811)
Hullfish, H.G., & Smith, P.G. (1968). *Reflective thinking: The method of education*. New York: Dodd, Mead.
Kim, J., & Mueller, C.W. (1978). *Introduction to factor analysis*. Beverly Hills, CA: Sage.
Loehlin, J.C. (1987). *Latent trait models*. Hillsdale, NJ: Erlbaum.
McDonald, R.P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
McPeck, J. (1981). *Critical thinking and education*. New York: St. Martin's.
Norris, S.P. (1986). *Test on appraising observations (constructed-response version)*. St. John's, NF: Memorial University of Newfoundland.
Norris, S.P. (1992). A demonstration of the use of verbal reports of thinking in multiple-choice critical thinking test design. *Alberta Journal of Educational Research, 38*, 155-176.
Norris, S.P., & Ennis, R.H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest.
Norris, S.P., & King, R. (1983). *Test on appraising observations (multiple-choice version)*. St. John's, NF: Memorial University of Newfoundland.
Norris, S.P., & Ryan, J. (1989). *Norris-Ryan argument analysis test*. St. John's, NF: Memorial University of Newfoundland.
Norris, S.P., & Ryan, J. (1992). *Test of inductive reasoning principles*. St. John's, NF: Memorial University of Newfoundland.
Perkins, D.N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly, 39*, 1-21.
Phillips, L.M. (1989). *Test of inference ability in reading comprehension (constructed-response format)*. St. John's, NF: Memorial University of Newfoundland.
Phillips, L.M., & Patterson, C.C. (1987). *Test of inference ability in reading comprehension (multiple-choice format)*. St. John's, NF: Memorial University of Newfoundland.
Resnick, L.B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.
Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. London: Routledge.
Smith, B.O. (1953). The improvement of critical thinking. *Progressive Education, 30*(5), 129-134.
Steiger, J.H. (1989). *Causal modeling*. Evanston, IL: SYSTAT.
Traub, R.E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. In R.E. Bennett & W.C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 29-44). Hillsdale, NJ: Erlbaum.
Traub, R.E., & Fisher, C.W. (1977). On the equivalence of constructed-response and multiple-choice tests. *Applied Psychological Measurement, 1*, 355-369.
Traub, R.E., & MacRury, K. (1990). Antwort-auswahl vs. freie-antwort-aufgaben bei Lernervolks-tests [Multiple-choice vs. free-response in the testing of scholastic achievement]. In K. Ingenkamp & R.S. Jäger (Eds.), *Tests und trends 8: Jahrbuchder pädagogischen diagnostik* (pp. 128-159). Weinheim, Germany: Beltz Verlag.
Watson, G., & Glaser, E.M. (1980). *Watson-Glaser critical thinking appraisal*. Cleveland, OH: Psychological Corporation.
Whimbey, A. (1985). You don't need a special "reasoning" test to implement and evaluate reasoning training. *Educational Leadership, 43*(2), 37-39.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 407-420*

José L. da Costa
University of Alberta

Teacher Collaboration: A Comparison of Four Strategies

Data for eight constructs were collected from 26 teachers and their 401 students in two public school districts in the British Columbia Lower Mainland. The question was: What teacher and pupil characteristics differentiate the four teacher collaboration strategies consisting of collaborative consultation, collaborative consultation with team teaching, collaborative consultation with no classroom observation, and collegial consultation with no classroom observation? Data collected for the eight constructs were analyzed using multivariate analysis of variance (MANOVA) followed by post-hoc discriminant analysis (DA). Results of the MANOVA suggest that differences did exist among the four collaboration groups. The post-hoc DA indicates that the groups differ on a dimension anchored by personal teaching efficacy and pupil achievement at one end and pupil attitude at the other. It is concluded that classroom observation is an essential element of the teacher collaboration process if increased pupil achievement and increased teacher efficacy are desired outcomes.

Des données de huit modèles théoriques ont été receuillies de 26 enseignant(e)s et leurs 401 élèves de deux districts scolaires au sud de la Colombie-Britannique. La question posée était qu'elles caractéristiques chez les enseignant(e)s et les élèves différencient les quatre stratégies de collaboration professionnelle suivantes: la consultation collaborative, la consultation collaborative dans un contexte d'enseignement en équipe, la consultation collaborative sans observation directe en salle de classe, et la consultation collégiale n'ayant aucune observation en salle de classe. Les données recueillies pour les huit modèles théoriques ont été analysées selon l'analyse de variantes multivariées (AVAMUL) suivie d'une analyse destinatrice post-hoc (AD). Les résultats de cette première analyse (AVAMUL) suggèrent qu'il existe en effet une différence entre les quatre différents groupes collaboratifs. La deuxième analyse post-hoc (AD) révèle que les groupes diffèrent sur une dimension ancrée d'une part sur l'efficacité d'enseignement personnel et le succès académique des élèves et de l'autre part, sur l'attitude des élèves. Il en déterminé que l'observation en salle de classe est un élément essentiel du processus de la collaboration professionnelle si le succès amélioré des élèves et l'efficacité accrue chez les enseignant(e)s sont les résultats désirés.

Over the past two decades educational researchers have conducted many studies investigating the effects on teachers and pupils of various teacher development approaches that emphasize teacher growth (Donovan, Sousa, & Walberg, 1987; Showers, 1985; Smith, 1989; Stallings, 1985). Many of these once-promising teacher development approaches have fallen into disfavor with both the research community and the teachers themselves (Slavin, 1986; Smith & Acheson, 1991; Stallings & Krasavage, 1986). The failure of these teacher development approaches has been attributed to various factors, one of which is lack of teacher commitment to the process due to conflicts between teachers' own norms and values and those imposed externally by the approach taken (Hargreaves & Fullan, 1992). Grimmett, Rostad, and Ford (1992) echo this sentiment:

José da Costa is an assistant professor in the Department of Educational Policy Studies.

> Externally mandated change typically has a cataclysmic effect on teachers' morale, resulting in a strong sense of dependency. Teachers often feel overwhelmed by the new expectations when their actions are continually shaped by the directives of others. There is an accompanying sense of helplessness and powerlessness when heightened expectations appear to be beyond reach. (pp. 185-186)

Consequently, of interest in the study reported in this article was a comparison of teacher consultation approaches the goal of which was to permit teachers to make sense of their classroom behaviors through their own values and norms. This interest, however, was not limited to the effects of the process of consultation on teachers but also examined the effects of the process of teacher consultation on students.

Garman (1986) contended that "ultimately the reason teachers and clinical supervisors work together is in order to enhance practice and to make education better for students" (p. 19). Similarly, Fullan and Hargreaves (1991) believed that improving teachers' pedagogic skills and school working conditions in turn would lead to student improvement. In other words, when teachers collaborate, their students should experience positive change in terms of achievement, attitude, and behavior. Presently, however, these links are only conjectures; empirical evidence is lacking (Acheson & Gall, 1992; Greene, 1987; Wildman & Niles, 1987). At a practical level the purpose of this study was to provide some initial direction to educators in developing teacher collaboration programs that benefit teachers as well as their pupils. Given the myriad possibilities in which teachers can work with other teachers, this study sought to collect empirical evidence to address the question: What teacher and pupil characteristics differentiate the four teacher collaboration strategies consisting of collaborative consultation (CC), collaborative consultation with team teaching (CCTT), collaborative consultation with no classroom observation (CCNO), and collegial consultation with no classroom observation (CoNO)? These four groups were chosen for the simple and practical reason of accessibility. Descriptions of the characteristics defining each group are found in Appendix A.

Review of the Literature

The literature suggests that many variables may be linked to teacher collaboration and its effects on teachers and pupils. This brief review explores those teacher and pupil variables that appear to be related to teacher collaboration.

Teacher growth through the use of teacher collaboration techniques is predicated on the establishment of teacher trust for the teaching partner (Acheson & Gall, 1992; Cogan, 1973; Goldhammer, Anderson, & Krajewski, 1980; Grimmett & Erickson, 1988; Hargreaves & Dawe, 1990; Lovell & Wiles, 1983; Sergiovanni & Starratt, 1993). This approach leads to the expectation that teacher trust for the teaching partner is influenced by the supervisory mode preferred by the teaching partner. Using Glickman's (1990) terms, the teaching partner's preferred mode of supervisory interaction should not be directive, but rather nondirective or collaborative if the intent is to promote trust for the teaching partner.

A potentially positive outcome of teacher collaboration when the above conditions are met is an increase in general and personal teaching efficacy

(Ashton, Webb, & Doda, 1983; Cavers, 1988; Denham & Michael, 1981). Cruickshank and Applegate's (1981) conclusions suggest that increased levels of self-esteem and a positive outlook on teaching leads teachers to become more concerned with self-improvement. Presumably, as a consequence of becoming more concerned with self-improvement teachers change their classroom behavior. Increased general and personal teaching efficacy can be thought to give the teacher confidence that improvement can be achieved by trying new approaches to solving classroom problems (Bandura, 1978).

Many authors (Acheson & Gall, 1992; Fullan & Hargreaves, 1991; Little, 1987; Lovell & Wiles, 1983; Sergiovanni & Starratt, 1993) agree that if teachers can be encouraged to think critically about what they do in the classroom and then act on those thoughts, there should be an accompanying change in pupils' achievement, attitudes, and behavior. However, these linkages are theoretical and have not as yet been demonstrated empirically.

This review suggests that at least eight teacher and pupil constructs are of importance for the present investigation of how teacher collaboration affects teachers and pupils. These constructs include: (a) teacher trust for the teaching partner, (b) the teaching partner's supervisory beliefs, (c) general teacher efficacy, (d) personal teacher efficacy, (e) teacher classroom behavior, (f) pupil achievement, (g) pupil attitude, and (h) pupil behavior.

Method

The Participants

The study employed a pre- postmeasurement survey design with no researcher intervention. The sample consisted of 26 self-selected classes (26 teachers and 401 pupil volunteers) at the elementary level from two public school districts in the British Columbia Lower Mainland. All of these teachers had previously participated in the British Columbia Teachers' Federation Program for Quality Teaching, which provided instruction and practice in the use of peer supervision techniques for the purpose of professional development (Smith & Acheson, 1991).

Four consultation strategies—arrived at inductively—describe how the teacher volunteers worked with their teaching partners: (a) collaborative consultation between teacher dyads teaching in separate classrooms, (b) collaborative consultation between team teacher dyads teaching in one double sized classroom, (c) collaborative consultation without direct observation with dyads teaching in separate classrooms, and (d) collegial consultation without direct observation (descriptions of the nature of each group are found in the Appendix). To determine into which of the four groups teachers fitted best each was asked the following questions to determine the nature of the professional relationship between the teacher and teaching partner: (a) Will you collaborate with a teaching partner? (b) In what capacity is your teaching partner employed in the school district? (c) Will your teaching partner be observing your classroom teaching to collect data for conferencing purposes? and (d) Do you share one open classroom with your teaching partner?

The Measures

Data were collected for the eight constructs identified in the literature as relevant. The scales used to obtain measures of each of the eight variables

deemed important by the literature were, respectively, as follows: (a) Wheeless and Grotz's (1977) Individualized Trust Scale (ITS), (b) Glickman and Tamashiro's (1981) Supervisory Beliefs Inventory (SBI), (c) Gibson and Dembo's (1984) Teacher Efficacy Scale—general dimension (TESg), (d) Gibson and Dembo's (1984) Teacher Efficacy Scale—personal dimension (TESp), (e) Eash and Waxman's (1983) Our Class and Its Work (OCIW), (f) pupil achievement grades from district report cards (ACH), (g) Randhawa and Van Hesteren's (1983) School Attitude Scales for Children (SASC), and (h) pupil behavior grades (BEH) from district report cards (see Appendix for details regarding each of these measures).

Data collection from took place at two points during the school year. Premeasure data were collected one month after the beginning of the school year, and the postmeasure data were collected one month before the end of the school year. Identical scales were used for both sets of measurements.

Data Analyses

Because the sample consisted of self-selected volunteers, it was decided that the analyses of the data should be carried out in three phases: (a) the first phase compared the sample with other samples and populations on the basis of five variables, (b) the second phase compared the four groups with one another on the basis of the premeasures data, and (c) the third phase compared the four groups on the basis of the postmeasures data. In order to prevent inflation of statistical power, all analyses used the class as the unit of analysis. This conservative approach was chosen to reduce the possibility of making a type I error in the statistical analyses.

First, descriptive statistics for each variable were calculated to screen the data and gain a preliminary understanding of the data set. Second, group differences on the premeasures were sought. A multivariate analysis of variance (MANOVA) was conducted on all the premeasure variables simultaneously to determine if the four collaboration groups differed near the beginning of the school year. Third, group differences on the postmeasures were sought using MANOVA. This was followed by post-hoc discriminant analysis (DA) to determine how the collaboration groups differed.

Limitations

The limitations of this study fall into two basic areas: (a) the use of a volunteer sample, and (b) the method of data analysis. Each of these is examined in turn.

This study drew data from a sufficiently large sample that, had it been randomly selected, the generalizability of the results to the population of teachers working in the suburban area surrounding Vancouver would not have been problematic. However, because the sample consisted entirely of teacher and pupil volunteers the results are not truly generalizable beyond the population of teachers and pupils who would volunteer for such a study. In other words, it is recognized that the variables associated with teacher collaboration are themselves affected by the degree to which the participants wish to participate. This fact is less problematic than it initially might sound. In discussing how to best implement professional development programs utilizing teacher collaboration, Hargreaves and Fullan (1992) and Greene (1992) both place a great deal of emphasis on the desirability of having teachers choose

whether to participate and, if so, how to participate in such programs. Presumably, the types of data gathered from the volunteer sample in this study would more closely resemble those cases in which teachers decide whether to participate than those cases in which teachers are told that they will collaborate with colleagues.

With respect to the limitations of the method of analysis, Patton (1990) stated *"there are no perfect research designs"* (original emphasis, p. 162). The method employed here is no exception. Analyses making use of MANOVA are more difficult to interpret than corresponding univariate analyses. However, the disadvantage of this limitation is greatly outweighed by the advantage of being able to simultaneously make sense of the differences among several groups using several constructs by taking into account the covariation of the variables of interest (Tabachnick & Fidell, 1989).

Findings

From the preceding analyses three categories of findings emerge, which are best classified as relating to: (a) the sample, (b) the premeasures, and (c) the postmeasures. Each is addressed below.

The Sample

On the basis of teacher experience, female-to-male ratio, teacher efficacy scores (TESg and TESp), and OCIW scores of teachers and their pupils appear to be reasonably representative of other teachers and pupils in other British Columbia school districts. Using a chi-square goodness of fit test, the teaching experience of the sample teachers was found not to be significantly different ($\alpha = .05$) from the teaching experience of the elementary teachers in the two districts from which the sample was drawn. Similarly, the female-to-male teacher ratio of 2:1 did not differ significantly ($\alpha = .05$) from the British Columbia provincial female-to-male elementary teacher ratio. In the vast majority of school districts, female elementary teachers greatly outnumber male teachers. Comparing the TESg, TESp, and OCIW mean scores obtained from the present sample with comparable mean scores from previous studies, this sample was seen to be similar to the samples (see Table 1) by Anderson, Greene, and Loewen (1987), Cavers (1988), Tracz and Gibson (1986), Grimmett and Crehan (E.P. Crehan, personal communication, April 22, 1992), and Housego (1992).

The Premeasure Analysis

A MANOVA was conducted to determine if any differences existed among the groups near the beginning of the school year. This MANOVA compared the four collaboration groups simultaneously on the basis of the eight constructs of interest.

After taking all the premeasure variables together to account for the covariation among the variables, the MANOVA revealed no significant differences among the four collaboration groups at the .05 level (Wilks lambda = .331, $F = .853$, $df = 24/44$, $p = .656$). Having established that the groups were probably similar at the beginning of the school year, the decision was made that the postmeasure group data did not need to be adjusted (i.e., through MANCOVA) to compensate for any initial differences among the groups.

Table 1
Comparisons of TES and OCIW Means Between the Present Study and Previous Studies

Study			Teacher Efficacy Personal	Teacher Efficacy General		OCIW Total
		n	mean	mean	n	mean
Anderson et al. (1987)		24	4.77	3.56	584	2.70
Cavers (1988)		339	4.64	3.44		
Tracz & Gibson (1986)		14	4.86	3.65		
Grimmett & Crehan (n.d.)	PRE	93	4.29	3.22		
	POST	78	4.69	3.20		
Housego (1992)	BASE	177	3.97	3.76		
	T1	129	4.38	3.56		
	T2	121	4.50	3.41		
	T3	123	4.47	3.47		
da Costa (1993)	PRE	26	4.56	3.75	401	2.68
	POST	26	4.53	3.62	401	2.68

The Postmeasure Analysis

The results of a MANOVA suggest that differences do exist at the .05 alpha level among the collaboration groups after taking all the postmeasure variables together to account for the covariation among the variables (Wilks lambda = .139, $F = 1.79$, $df = 24/44$, $p = .046$). To find the nature of these differences a post-hoc DA was conducted.

Using the eight postmeasure variables as potential predictors, DA was carried out. At the .05 alpha level, only the first discriminant function was retained ($\chi^2(24) = 37.51$, $p = .039$). Using structure coefficients, with loadings of |.25| or greater, a dimension with pupil attitude (–.44) at one end and personal teaching efficacy (.27) along with pupil achievement (.39) at the other end is described. As can be seen in Figure 1, plotting the discriminant means (CC mean = 1.89, CCTT mean = 0.81, CCNO mean = –1.59, and CoNO mean = 0.00), these findings were obtained: (a) CC group differed most from the other three groups; (b) CCTT group and the CoNO groups were most similar to each other, but they differed from both the CC and the CCNO groups; and (c) the CCNO group differed from all other groups.

The CC group seemed to stand out when compared with other collaboration groups in the present study. Teachers in the CC group exhibited more personal teaching efficacy, and pupils in this group had higher achievement than the teachers and pupils, respectively, in the other collaboration groups. However, pupils in the CC group also tended to have more negative attitudes toward these attitudinal objects addressed in the SASC: themselves, peers, their teacher, school, learning in general, language arts, social studies, and math. Also distinctive in comparison with the other collaboration groups was the CCNO group. Teachers in this group tended to have lower personal teaching

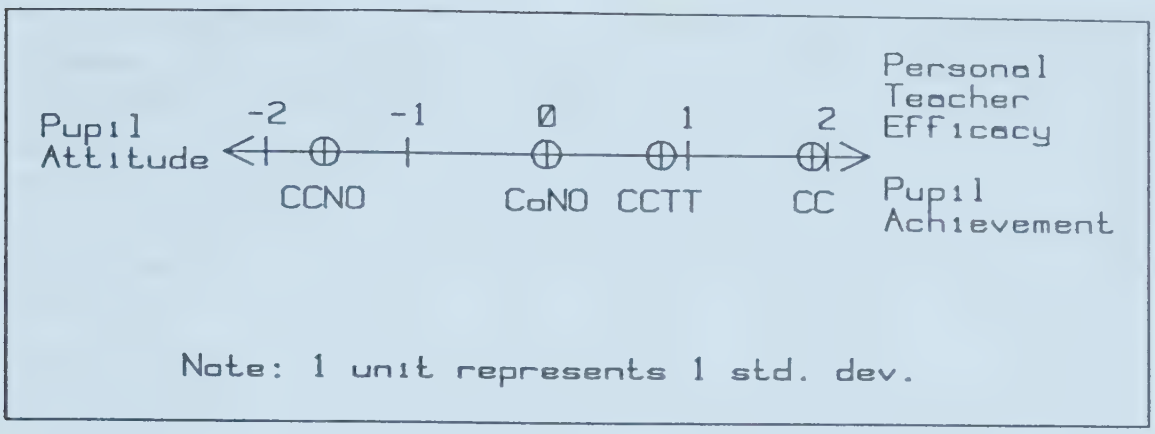


Figure 1. The single bipolar dimension on which the four collaboration groups discriminated.

efficacy and lower pupil achievement, but the same pupils generally had more positive attitudes toward the attitudinal objects addressed by the SASC.

Discussion

It is noteworthy that the post-hoc DA resulted in at least one discriminant function describing a bipolar underlying continuum composed of personal teaching efficacy and pupil achievement at one end of the scale and pupil attitudes at the opposite end. The expectation as a result of the literature reviewed (Acheson & Gall, 1992; Little, 1987) would have been to observe pupil attitudes and pupil achievement anchor at one end of the continuum rather than at opposite ends of the same continuum as found here. This suggests that high achieving pupils generally have more negative attitudes toward attitudinal objects associated with school—at least toward those measured by the SASC.

A second point regarding this continuum involves the relationship between personal teaching efficacy and pupil achievement. Several studies suggest that a reciprocal relationship exists between teaching efficacy and pupil achievement (Armor et al., 1976; Ashton, 1985; Berman, McLaughlin, Bass, Pauly, & Zellman, 1977). Cavers (1988), however, found no evidence of such a relationship. Ashton and Webb (1986) concluded that their findings "strongly support the hypothesis that teachers' sense of efficacy is related to pupil achievement" (p. 139). The present study suggests that (a) a positive relationship does exist between personal teaching efficacy and pupil achievement as suggested by Anderson et al. (1987), and (b) a negative relationship exists between personal teaching efficacy and pupil attitudes toward school-related attitudinal objects. Furthermore, Canon and Simpson's (1985) and Randhawa and Van Hesteren's (1982) conclusion that attitude and achievement are not significantly related is also supported by the present study.

A possible explanation for the grouped data relationships among personal teaching efficacy, pupil achievement, and pupil attitudes may be found by first recalling the definition of personal teaching efficacy and then examining the nine items related to this efficacy dimension on Gibson and Dembo's (1984) TES. Personal teaching efficacy is described as the belief that the individual teacher has the skills and abilities to bring about pupil learning. This definition is reflected in the personal teaching efficacy items of the TES by the inclusion of

the phrase indicating how the teacher would react to various difficulties. Examples of phrases indicating how the teacher would react to pupil problems include: item 1, "I exerted a little extra effort"; item 5, "I am usually able"; item 6, "I found better ways"; item 13, "I know some techniques." Teachers with high personal teaching efficacy may be more apt than teachers with lower personal teaching efficacy to try various techniques or strategies to increase individual pupil achievement. These techniques or strategies may succeed in increasing pupil achievement; however, pupils may no longer feel they have ownership of the learning process because the teacher has control of it. It may be speculated that this loss of control of the learning process on the part of the pupil may result in more negative pupil attitudes toward school-related attitudinal objects (B.E.J. Housego, personal communication, March 18, 1993).

The general trend of positioning the various group means relative to one another in the discriminant space comprising personal teaching efficacy and pupil achievement at one extreme and pupil attitudes at the other (as seen in Figure 1) was for the most part expected on the basis of what had been indicated in the literature. The expectation was that if groups were going to cluster they would do so according to the following: (a) the collaborative consultation groups would separate from the collegial consultation group, and (b) the groups utilizing direct classroom observation would separate from those groups not utilizing direct classroom observation. As seen in Figure 1 the trend is for the groups using direct classroom observation to cluster as expected (i.e., the collaborative consultation group and the collaborative consultation in a team teaching situation group are positioned side by side in discriminant space). Teachers basing their collaboration conferences on classroom observational data gathered by a peer were better able to effect changes to improve their pupils' levels of achievement. An offshoot of this seems to be increased personal teaching efficacy for the individual teacher. Also as expected, the groups not making use of classroom observation on which to base their follow-up conferences tended to fall on the continuum away from the end anchored by high levels of personal teaching efficacy and pupil achievement.

The collaborative consultation with no observation fell so much farther to the left of the collegial consultation group on this continuum, indicating even lower levels of personal teaching efficacy and pupil achievement. This was surprising because those teachers involved in collegial consultation were not conferring with a peer, but were working in nonreciprocal arrangements with "experts" (i.e., administrators, master teachers) who were in a position to tell the teachers what needed to be done in order to address problems in the teachers' classrooms.

Conclusions

Given the exploratory nature of the present study, the conclusions reached and the recommendations made need to be viewed cautiously. This study is useful from a practical perspective because it suggests some ways for organizing teacher development through collaboration programs and from a theoretical perspective because it provides a first step in obtaining empirical evidence to show how teachers working with other teachers are affected and how their pupils are affected.

The Importance of Classroom Observation

Classroom observation on the part of the teaching partner is an essential element of collaborative consultation if the aim is to increase personal teaching efficacy and pupil achievement. This is what can be expected when collaborative consultation is conducted using Garman's (1986) model of reflection on action which is based on the contextualization by the teacher of "stable" data in light of educational theory through interpretation, explanation, and evaluation.

Collaboration not Based on Classroom Observation

As noted above, conferences based on the data collected from direct classroom observation provide better results when the desire of the teacher is to improve pupil achievement and increase personal teaching efficacy. However, this study suggests that situations that preclude the use of data collection from direct classroom observation are better handled using a nonreciprocal consultation model between a teacher and a pedagogical expert (i.e., administrator, master teacher). Collaborative consultation without direct classroom observation on the part of the teaching partner was the least effective method of the four studied in promoting pupil achievement and personal teacher efficacy.

Relationship Among Personal Teacher Efficacy, Pupil Achievement, and Pupil Attitude

The positive relationship found between personal teacher efficacy and pupil achievement, the negative relationship found between personal teacher efficacy and pupil attitudes, and the negative relationship found between pupil attitudes and achievement seem to contradict some intuitive notions of how these constructs are related. Given this finding and the work of Anderson et al. (1987) and of Ashton and Webb (1986), the conclusion that increases in personal teaching efficacy correspond to increases in pupil achievement seems reasonable. Adding to this the work of Canon and Simpson (1985) and Randhawa and Van Hesteren (1983), who speculated that achievement and attitude were not directly related, it is reasonable to conclude that achievement and attitude are not similarly related to personal teaching efficacy. Rather, increases in personal teaching efficacy correspond to changes in pupil attitude. Furthermore, the mechanism that relates personal teaching efficacy to pupil achievement and attitude is not at this time clearly evident.

Recommendations

In addition for a call to replicate this study, three recommendations follow from the three conclusions drawn above. The first two address practice, the third theory. The recommendations for practice are made under these assumptions (a) teacher participation in a teacher collaboration program is not mandated by district personnel, and (b) teachers select the pedagogic areas in which they wish to change or improve.

Recommendation 1. Teacher collaboration programs in which classroom observation by the teaching partner is an integral part of the design should be encouraged because they seem to be more positively associated with increased personal teacher efficacy and higher pupil achievement than nonobservation collaboration approaches. Typically, teaching partners must be provided re-

lease time from their own classrooms to observe the teacher and his or her class so that the teacher can be provided with what Garman (1984) calls stable data.

Recommendation 2. In situations where teacher collaboration programs are implemented that do not include classroom observation (i.e., the data discussed during conferencing are recalled by the teacher) by the teaching partner as an integral part of the design, teachers should discuss their teaching practices with an expert (i.e., administrator, master teacher) in a nonreciprocal relationship. This is not to say that implementing teacher collaboration programs that do not utilize classroom observation for data collection are recommended, but it is acknowledged that in many circumstances the resources required for this type of program are simply not available.

Recommendation 3. To help understand the mechanism by which personal teaching efficacy, pupil achievement, and pupil attitude are related, a factor analysis of the SASC pupil attitude data should be conducted to determine the nature of the dimensions that underlie the attitude construct. Intuitively, the expectation is that when pupil attitudes change, according to any one of the SASC subscales used in the present study, pupil favorableness toward or against other psychological objects of interest may also change.

Acknowledgements

The feedback provided by Dr. E.A. Holdaway and the reviewers of the draft of this manuscript is gratefully acknowledged.

References

Acheson, K.A., & Gall, M.D. (1992). *Techniques in the clinical supervision of teachers: Preservice and inservice applications* (3rd ed.). New York: Longman.

Anderson, R., Greene, M., & Loewen, P. (1987). *Relationships among teachers' and students' thinking skills, sense of efficacy and student achievement*. Unpublished manuscript, University of Lethbridge.

Armor, D., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascall, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools*. Santa Monica, CA: Rand.

Ashton, P.J. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), *Research on motivation in education, Volume 2: The classroom milieu* (pp. 141-169). Toronto, ON: Academic Press.

Ashton, P.J., & Webb, R.B. (1986). *Making a difference*. New York: Longman.

Ashton, P.J., Webb, R., & Doda, N. (1983). *A study of a teacher's sense of efficacy, final report. Executive summary*. Gainsville, FL: Florida University.

Bandura, A. (1978). The self system in reciprocal determinism. *American Psychologist, 33*, 344-358.

Berman.P., McLaughlin,M., Bass, G., Pauly, E., & Zellman, G. (1977). *Federal programs supporting educational change. Vol. VII: Factors affecting implementation and continuation*. Santa Monica, CA: Rand.

Canon, R.K., & Simpson, R.D. (1985). Relationships among attitude, motivation, and achievement of ability grouped, seventh-grade, life science students. *Science Education, 69*(2), 121-138.

Cavers, L. (1988). *Teacher efficacy: Its relationship to school level organizational conditions and teacher demographic characteristics*. Unpublished doctoral dissertation, University of British Columbia.

Cogan, M.L. (1973). *Clinical supervision*. Boston, MA: Houghton Mifflin.

Cruickshank, D.R., & Applegate, J.H. (1981). Reflective teaching as a strategy for teacher growth. *Educational Leadership, 38*(7), 553-554.

Denham, C., & Michael, J. (1981). Teacher sense of efficacy: A definition of the construct and a model for future research. *Educational Research Quarterly, 6*(1), 39-63.

Donovan, J.F., Sousa, D.A., & Walberg, H.J. (1987). The impact of staff development on implementation and student achievement. *Journal of Educational Research, 80*, 348-351.

Eash, M.J., & Waxman, H.C. (1983). Utilizing students' perceptions and context variables to analyze effect teaching: A process-product investigation. *Journal of Educational Research, 76*(6), 321-324.

Fullan, M., & Hargreaves, A. (1991). *What's worth fighting for? Working together for your school.* Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 342 128)

Garman, N.B. (1984, April). *Stable data and clinical supervision.* Paper presented at the annual conference of the Association of Teacher Educators, New Orleans.

Garman, N.B. (1986). Reflection, the heart of clinical supervision: A modern rationale for professional practice. *Journal of Curriculum and Supervision, 2*(1), 1-24.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 565-582.

Glickman, C.D. (1990). *Supervision of instruction: A developmental approach* (2nd ed.). New York: Allyn and Bacon.

Glickman, C.D., & Tamashiro, R.T. (1981). The supervisory beliefs inventory. In C.D. Glickman (Ed.), *Developmental supervision: Alternative practices for helping teachers to improve instruction* (pp. 88-91). Alexandria, VA: Association for Supervision and Curriculum Development.

Goldhammer, R., Anderson, R.H., & Krajewski, R.J. (1980). *Special methods for the supervision of teachers* (2nd ed.). New York: Holt.

Greene, M. (1987). *Landscapes of learning.* New York: Teachers College Press.

Greene, M. (1992). Teacher supervision as professional development: Does it work? *Journal of Curriculum and Supervision, 7*(2), 131-148.

Grimmett, P.P., & Crehan, E.P. (n.d.). *Collaborative consultation project.* Unpublished manuscript, University of British Columbia.

Grimmett, P.P., & Erickson, G.L. (1988). *Reflection in teacher education.* Vancouver, BC: University of British Columbia, Pacific Educational Press.

Grimmett, P.P., Rostad, O.P., & Ford,B. (1992). The transformation of supervision. In C. Glickman (Ed.), *Supervision in transition, 1992 yearbook of the Association for Supervision and Curriculum Development* (pp. 185-202). Alexandria, VA: Association for Supervision and Curriculum Development.

Hargreaves, A., & Dawe, R. (1990). Paths of professional development: Contrived collegiality, collaborative culture, and the case of peer coaching. *Teaching and Teacher Education, 6*(3), 227-241.

Hargreaves, A., & Fullan, M. (1992). Shortcomings of traditional approaches to teacher development. *Keys to Teacher Development, 1*(1), 3-4.

Housego, B.E.J. (1992). Monitoring student teachers' feelings of preparedness to teach, personal teaching efficacy, and teaching efficacy in a new secondary teacher education program. *Alberta Journal of Educational Research, 38, 49-64.*

Little, J. (1987). Teachers as colleagues. In V.R. Koehler (Ed.), *Educators' handbook: A research perspective* (pp. 491-518). New York: Longman.

Lovell, J.T., & Wiles, K. (1983). *Supervision for better schools* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Patton, M.Q. (1990). *Qualitative evaluation and research methods.* Newbury Park, CA: Sage.

Randhawa, B., & Van Hesteren, F. (1982, June). *The development and validation of a hybrid semantic differential measure of school related attitudes in young children.* Paper presented at the annual meeting of the Canadian Psychological Association, Montreal. (ERIC Document Reproduction Service No. ED 223 691)

Randhawa, B., Van Hesteren, F. (1983). *School-related attitudes and their redundancy with academic achievement.* Paper presented at the Learned Societies Conference of the Canadian Society for the Study of Education, Vancouver. (ERIC Document Reproduction Service No. ED 235 227)

Sergiovanni, T.J., & Starratt, R.J. (1993). *Supervision: Human perspectives* (5th ed.). New York: McGraw Hill.

Showers, B. (1985). Teachers coaching teachers. *Educational Leadership, 42*(7), 43-48.

Slavin, R.E. (1986). The Napa evaluation of Madeline Hunter's ITIP: Lessons learned. *Elementary School Journal, 87*(2), 165-171.

Smith, N.S. (1989). *Peer consultation as staff development. A study in BC Schools.* Unpublished doctoral dissertation, University of Oregon.

Smith, N.S., & Acheson, K.A. (1991). Peer consultation: An analysis of several types of programs. *Oregon School Study Council, 34*(6), 1-29.

Stallings, J. (1985). A study of implementation of Madeline Hunter's model and its effects on students. *Journal of Educational Research, 78*(6), 325-337.

Stallings, J., & Krasavage, E. (1986). The effects of instruction based on the Madeline Hunter model on student's achievement: Findings from a follow-through project. *Elementary School Journal, 87*(2), 571-588.

Tabachnick, B.G., & Fidell, L.S. (1989). *Using multivariate statistics* (2nd ed.). New York: Harper and Row.

Tracz, S., & Gibson, M. (1986, November). *Effects of efficacy on academic achievement*. Paper presented at the California Educational Research Association Annual Meeting, Marina del Rey. (Eric Document Reproduction Service No. ED 281 853)

Wheeless, L.R., & Grotz, J. (1977). The measurement of trust and its relationship to self disclosure. *Human Communication Research, 3*(3), 250-257.

Wildman, T.M., & Niles, J.A. (1987). Reflective teachers: Tensions between abstractions and realities. *Journal of Teacher Education, 38*(4), 25-31.

Appendix A
Description of Terms

Clinical supervision. This refers to the partnership described by Cogan (1973) and Goldhammer et al. (1980) between supervisor and teacher that uses classroom data as the basis for subsequent analyses the purpose of which is to improve the teacher's classroom practices in ways that make sense to the teacher. This process is intended to be for formative purposes only. The supervisor's job is not to identify what is "right" or "wrong" with the teacher's teaching but to help the teacher identify appropriate goals for improvement. The teacher's role is to decide the focus of the clinical supervision process and the direction in which it will proceed.

Collaborative consultation. This also refers to a process intended to facilitate teacher development using the principles of Cogan's (1973) clinical supervision. Underpinning this process, however, is a nonhierarchical relationship between a teacher and a teaching partner characterized by mutual trust and respect that is presumed to provide a supportive environment in which the teacher can evaluate previous teaching strategies as well as implement and evaluate new strategies. It is a relationship of equals in which both partners wish to engage (E.P. Crehan, personal communication, February 15, 1990).

Collaborative consultation without direct observation. This is similar to collaborative consultation as described above except that the classroom observation phase, as described by Cogan (1973), used to collect "objective" data is not implemented; instead, data for conferencing are obtained from the teacher's recollection of past events.

Collegial consultation. This refers to a process intended to facilitate teacher development using the principles of Cogan's (1973) clinical supervision. This is a professional relationship between a teacher and another individual in a school or school district (e.g., principal, master teacher, curriculum specialist, etc.). The relationship is not reciprocal: the two people forming the dyad do not exchange roles.

Collegial consultation without direct observation. This similar to collegial consultation as defined above except that the classroom observation phase, as described by Cogan (1973), used to collect "objective" data is not implemented; instead, data for conferencing are obtained from the teacher's recollection of past events.

Individualized Trust Scale. Wheeless and Grotz's (1977) ITS is a unidimensional scale measuring one's trust for another specific individual. The scale consists of 15 semantic differential-type, 7-interval items with the positive-negative

polarities randomly ordered to avoid response bias. A split-half reliability coefficient of 0.92 was reported when the scale was administered to 100 teacher and their spouses or eldest children (n = 261). The authors report that "the instrument had predictive validity ... for general use as an alternate means of measuring trust" (p. 256).

Our Class and Its Work. Eash and Waxman's (1983) OCIW instrument was used for gathering student perceptions of teacher classroom behavior. The instrument consists of 40 items describing teaching behaviors that form eight Likert-type subscales, namely, didactic instruction, enthusiasm, feedback, instructional time, opportunity to learn, pacing, structuring comments, and task orientation. High reliabilities for the eight scales are reported; after administering the scale to 762 pupils in 36 grade 4 and 6 classes in a large American city school district Cronbach alphas range from 0.84 to 0.92. To assess construct validity, an examination of the relationships among OCIW scores, student achievement, and principal ratings of the teachers obtained from the sample described showed that teachers who obtained higher scores on the OCIW scale also "were rated higher by the principal in his yearly evaluations" (p. 4).

Pupil achievement grades. This study relied entirely on the teachers' anecdotal and letter graded report cards for establishing levels of pupil achievement. Achievement data from teachers' anecdotal notes on report cards of pupil progress were used to check the veracity of the teachers' assigned behavior grades. Two report cards of pupil progress were used in this study: (a) November 1991—premeasure, and (b) June 1992—postmeasure. Pupil report card data (letter grades) regarding achievement (e.g., language arts, math, science, and social studies were equally weighted to determine an overall achievement score) were coded using a four-point grade point average (GPA) type scale (A+, A, A– were assigned a 3, B+, B, B– were assigned a 2, C+, C, C– were assigned a 1, D+, D, D–, F were assigned a 0).

Pupil behavior grades. Pupil behavior data were collected from the November 1991—premeasure—and the June 1992—postmeasure report cards. Behavior data from teachers' anecdotal notes on report cards of pupil progress were used to check the veracity of the teachers' assigned behavior grades. Categorical data were converted to numerical data using the following schema: (a) good or excellent—assigned a 2, (b) satisfactory—assigned a 1, and (c) poor or needs improvement—assigned a 0. These behavior grades were then averaged to obtain an overall class behavior grade.

School Attitudes Scale for Children. The SASC consists of a series of "semantic differential scales that tap the following school-related dimensions: School, Teachers, Arithmetic, Science, Social Studies, Language Arts, Music, Drama, French, Art, Dance, Religion, Health, and Physical Education" (Randhawa & Van Hesteren, 1982, p. 6). The authors reported reliability coefficients ranging from 0.89 to 0.97 for the scales when administered to 99 pupils in grades 3-6. The authors also report that the scales exhibit a high degree of construct validity when analyzed using a "multitrait-multimethod convergent and discriminant validation methodology" (p. 11). For the present study four of the original SASC scales were used: school, social studies, language arts, and arithmetic. In addition to these, four other SASC scales—yourself, your

classmates, your regular teacher, and learning in general—were created to obtain measures of what the literature suggests are important attitudinal objects.

Supervisory Beliefs Inventory. The SBI (Glickman, 1990) is used to estimate the proportion of preferred modes of interaction—directive, collaborative, and nondirective—between a teacher and a teaching partner. The SBI consists of 15 items that when scored indicate the approximate proportion of preference for each of the interaction approaches. The literature reveals no reliability measures for the SBI. With respect to validity, the author states that "the instrument has been field-tested six times with ninety supervisors and supervisor trainees. Response between the options indicated 'good'" item discrimination" (p. 91).

Teacher Efficacy Scale. Gibson and Dembo's (1984) Likert-type teacher efficacy scale contains two factors: general teaching efficacy, the belief that teachers as a collective can influence pupil learning; and personal teaching efficacy, the belief that the individual teacher can influence pupil learning. For this study the differentiation between general teaching efficacy and personal teaching efficacy was maintained. The authors report the following Cronbach's alpha coefficients for the 16-item teacher efficacy scale: 0.78 for the personal teaching dimension, 0.75 for the general teaching dimension. These findings are supported by Anderson, Greene, and Loewen (1987), who reported similar findings.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 421-434*

Judith Golec

John Gartrell
University of Alberta

and

William J. Sveinson
University of Victoria

University Performance of Nonmatriculated Admissions

Nonmatriculated admission (NMA) programs in universities have sought to promote access and equity while maintaining academic standards. These programs often assume that maturity is associated with improved motivation, interest, and decision making abilities and that life experience will provide the kind of contextual information that improves learning. These assumptions appear to be overly optimistic in their expectations. The academic performance of mature NMA admissions to the Faculty of Arts at the University of Alberta was in fact substantially poorer than the performance of matriculant admissions. The problem with this admission program may be attributable to abandonment of individual screening in favor of minimalist, formal admission criteria.

Les programmes d'admission aux universités pour les élèves non diplômés du cours secondaire 2^{e} cycle ont cherché à promouvoir l'accès et l'équité tout en maintenant certains standards académiques. Ces programmes présument souvent qu'à la maturité s'associe un accroissement de motivation, d'intérêt, et d'habiletés à prendre des décisions et que les expériences de vie fourniront l'information contextuelle nécessaire à l'amélioration de l'apprentissage. Ces suppositions sembleraient être trop optimistes dans leurs attentes. La performance académiques des élèves adultes non diplômés du cours secondaire 2^{e} cycle à la faculté des arts de l'Université de l'Alberta était en effet considérablement inférieure à celle des admissions des élèves diplômés du cours secondaires 2^{e} cycle. La difficulté que présente ce programme d'admission peut être attribuée à l'abandon des interviews individuelles des candidat(e)s en faveur des critères d'admission formels minimalisés.

Nonmatriculated admission (NMA) programs have been a part of Canadian universities' efforts to promote access and increase enrollments. More relaxed admission policies represent a response to equity issues, because mature admissions give people a "second chance." To the extent that universities become centers of adult retraining and learning over the whole life course, it may be

Judith Golec is an assistant professor in the Department of Sociology. She is currently studying the process of mutual aid in health-related social support groups.
John Gartrell is a professor in the Department of Sociology. He is reanalyzing the data used for *The Bell Curve* and beginning a study of development and the environment in Canada's Rocky Mountain communities.
William Sveinson is a research project coordinator at the Centre on Aging. He is presently providing management and technical support for a major study of independent community living among seniors.

seen as unfair to exclude a large number of people (Alberta Advanced Education, 1990; Anisef, Okihiro, & James, 1982; Vaselenak, 1970). The problem with flexible admission requirements is the maintenance of academic standards. Of course, the admission of nonmatriculated mature applicants assumes that previous failure can be attributed in large part to immaturity rather than inability or lack of academic preparation. The gaining of life experience that comes with greater maturity is expected to provide prospective students with the ability to overcome previous difficulties. For this reason nonmatriculated admission is sometimes called mature student admission.

Our focus is on standards of student performance. More specifically, do mature admissions to university perform as well as students who are admitted with full high school matriculation? The assumption of mature admission programs is that experience can compensate for prior poor performance. Experience is thought to provide increased commitment, motivation, interest, and decision making abilities, which make up for a lack of formal academic qualifications. Knowledge gained through life experience is thought to provide an enriched context for the interpretation of new knowledge (Eaton & West, 1980; Jackson, Small, & Zelmer, 1985; Perkins, 1971). Unfortunately, the efficacy of nonmatriculation admission policies is often just taken as a matter of faith. Professors may assume that mature students do better because they have had some mature students who were outstanding (overgeneralization), but there is remarkably little systematic, published evidence to support this article of faith.

In Canada, Darling (1985) documented the extent and diversity of mature admission programs. All 44 postsecondary institutions surveyed had special provisions to admit nonmatriculated applicants. Perkins (1971) concluded that mature admissions in Arts and Humanities at the University of Lethbridge were superior in their overall academic performance. "Maturity and motivation compensated for lack of past formal educational experience" (p. 111). Jackson et al. (1985) assert, on the basis of nonsystematic observation, that NMA performance at the University of Alberta was satisfactory. Wilson and Lapinski (1978) found similar results in a more systematic study conducted at Simon Fraser University, although a later unpublished report (Mature Student Entry Advisory Committee, 1983) comes to the opposite conclusion. There is some corroboration of the success of NMA programs from studies conducted in Australia (Barrett & Powell, 1980; Eaton & West, 1980) and Great Britain (Smithers & Griffin, 1986; Walker, 1975). More generally, a number of authors have documented a positive correlation between age and learning, even among adults (Johnson & Walberg, 1989; McLeish, 1962). As sparse as the empirical evidence appears to be, it seems to offer support for an organizational culture in universities that believes that mature students perform better than, or at least as well as, students with high school matriculation.

Admission policies of the Faculty of Arts at the University of Alberta provide an opportunity to empirically examine the academic performance of NMAs. In many respects the U of A's experience of relaxing normal admission requirements is not unusual. The earliest examples occurred during the 1940s. In the early 1940s the high school curriculum for the province as a whole was revised. However, not all secondary schools, especially rural schools, were able to offer the full high school curriculum. At the discretion of the dean of the

admitting faculty, some applicants with one or more matriculation deficiencies were admitted in the belief that returning to high school for upgrading was impractical. This is the same logic that was used a few years later to admit ex-servicemen whose education had been interrupted by World War II. It was also at this time that the University of Alberta began to offer matriculation (100 level) subjects as part of its own curriculum, thereby ensuring that deficiencies would be remedied during the first year of university.

Like most universities in Canada (Darling, 1985) it was not until the 1960s that the University of Alberta created a formal policy for admitting unqualified "adults" as part of a broader movement to increase access for students who had dropped out of formal education, the late bloomers, and the economically disadvantaged. The small numbers of applicants admitted to the Faculty of Arts (7 in 1965-1966; 31 in 1966-1967; and fewer than 65 in any year until 1973-1974) suggests a cautious application of the policy. In fact, unqualified applicants were required to undergo a rigorous screening process that included an initial interview with the Assistant Dean, a battery of tests administered by Student Counselling Services, an essay explaining the applicant's work experience and reason for desiring to attend university, followed by another interview with the Assistant Dean. Because decisions were based on the merits of the case, many students were advised to take remedial studies before reapplying. Interestingly, applicants lacking only one matriculation subject were regularly required to complete matriculation. Apparently, it was felt that taking one subject did not entail much hardship.

Between 1972-1973 and 1973-1974, enrollment of nonmatriculants jumped markedly (203) and in-house faculty reports document a concern that academic standards were being diluted. During the 1970s and early 1980s the nonmatriculated admission policy and procedures underwent many changes. Regrettably, because the experiments were not formally evaluated, it is not possible to state what impact the various changes may have produced. The minimum age was changed from 21 years to 24 and back to 21. For a short period NMAs were admitted "on probation." A minimum level of academic preparation was prescribed (grade 10) and later removed. Citizenship (Canadian or landed immigrant) and residency (Province of Alberta, Northwest Territories, or the Yukon) requirements were specified. Applicants lacking a matriculation level language other than English were admitted but on the condition that they clear the deficiency during the first year of the program; through practice this condition was extended with completion required sometime before graduation. Administering the battery of aptitude tests exacted a relatively high cost in organizational resources, and besides, very few applicants were rejected on these grounds alone. Screening applicants with a battery of psychological tests was reinterpreted as a barrier to access and the practice ceased. Similarly, the practice of requiring essays or letters of intent was also abandoned.

By the mid-1980s, the only admission criterion for NMAs besides age (21 years), citizenship, and residency was the successful completion of grade 12 English (English 30) or its equivalent (writing competence test or a university level English). If other academic work had been attempted in the three years preceding the application for mature admission, there could be no record of

failure. Apparently there is nothing unique about all this. According to the Darling (1985) survey of nonmatriculation admission policies, of the postsecondary institutions prescribing a minimum age (only 10 do not) the majority (23 out of 34) specify 21 years. Like the University of Alberta, roughly half (21 of 44) grant NMAs clear rather than probationary admission and most (33 of 44) use the same procedures to admit to full- or part-time status. Most (all but 6) have either ceased or never introduced the practice of screening NMAs with tests. Unlike the University of Alberta, about half still require a letter or résumé to evaluate the applicant's life experience and reason for entering a degree program.

Enrollment of NMAs increased. By 1985-1986 they comprised roughly 40% of freshman undergraduate admissions to the Faculty of Arts. In the absence of reliable data that describe enrollment trends for the NMA category, it is not possible to judge with certainty whether the Faculty of Arts at the University of Alberta is unique in this regard. Relying on impressions of the registrars surveyed, Darling (1985) estimates that NMA admissions make up fewer than 5% of all full-time undergraduate admissions and more for part-time admissions (some faculties in excess of 10%).

In the face of declining budgets, increasing demand for undergraduate admission, and a stable organizational resource base, the University of Alberta (like other universities) was forced to consider strategies for controlling undergraduate enrollments. In fact during the mid-1980s, the minimum admission average for matriculated applicants was raised from 65% to 70%, and discussion on the merits of establishing enrollment quotas in the undergraduate faculties was initiated. There is an obvious contradiction contained in these admission policies: increasing academic standards to reduce the numbers of fully qualified applicants while relaxing standards to increase the numbers of unqualified mature students. This puts a new turn on the issue of access and provides the social context for examining the academic performance of NMAs.

Our investigation was guided by a number of general propositions. First, based on previous findings and the admission policies for the Faculty of Arts and the University of Alberta, we would expect NMAs to perform at least as well as matriculation students. Second, to the extent that age reflects maturity, which in turn enhances performance (perhaps via experience and motivation), we expect age to have a positive effect on performance for both regular matriculation and NMA admissions. Third, we wish to examine the accuracy with which other university admission criteria (English 30 grade) predict academic performance. Finally, we extend our analysis to investigate how early performance in university predicts later performance, particularly after adjusting for the effects of age and admission criteria.

Research Design

Sample

All undergraduate students admitted (for the first time) to year 1 of a three- or four-year degree program in the Faculty of Arts between September 1985 and January 1986 were selected for study. This includes students applying under matriculation and nonmatriculation status.[1] The academic performance of the 1985-1986 study cohort was observed over a four-year period ending in August 1989.

The data for this analysis were collected entirely from student records (application form, academic transcripts) and therefore are limited to information routinely collected and held by the Registrar in electronic form. One advantage of using student records is the cost efficiency of available information. The use of records data also makes it likely that our analysis could be easily replicated at other universities or colleges. However, the University of Alberta student records were not designed for research use, and it was therefore necessary to make data transformations (described below under "measurement") in order to carry out the analysis. A more serious drawback to archival data is the limitation imposed by the paucity of information contained in such records. The records do not contain information on employment history, family commitments, or other variables that might help to shed light on questions of life experience, motivation, and maturity.[2]

Measurement

We selected grade point average (GPA) as our primary indicator of academic performance. GPA is constructed by dividing each student's total grade points by his or her total course weights taken in a specified time period. GPA has a number of advantages over other performance criteria, because it is commonly used by universities and the students themselves to evaluate student performance. As a composite of many professors' assessments of the students' performance, random errors in assessment are more likely to average out. At the University of Alberta course grades (and grade point averages) are measured on a 9-point scale (1-9). A score of 3 represents a failure, 4 is a marginal pass, 5 and 6 are satisfactory (roughly high and low Cs), 7 represents good performance (a B), 8 is very good or excellent (an A), and 9 is outstanding (an A+).

Sessional GPAs are calculated and printed on official transcripts at the University of Alberta. Traditionally, most students attended only winter session (September through April), and academic decisions for promotion and distinction continue to be based on the winter session GPA. In the case of matriculants, winter session GPA still tends to capture the yearly course load for most students. In the case of nonmatriculants, however, yearly course work often stretches across three sessions: winter, spring, and summer. To avoid distortion, we measured academic performance by computing an overall (four-year) GPA, that is, by summing grade points and course weights across the three sessions and four years under study. We were also interested in examining the effect of early (first-year) performance on later performance (years 2 through 4). Therefore, we computed a first-year GPA by summing grade points and course weights across the three sessions of 1985-1986 and a measure of later performance by computing a GPA for years 2 through 4 (summing across all sessions for 1986-1987, 1987-1988, and 1988-1989).

Other indicators of academic performance are also reported. These include graduation, first-class standing, receiving a scholarly award, and the requirement to withdraw. We compare the percentage of matriculants and nonmatriculants who graduated on or before spring convocation 1989, earned first-class standing (sessional GPA of 7.5 or higher on a full course load) in any winter session, and received at least one scholarship or prize in four years. Because nonmatriculants are more likely than matriculants to attend university as part-time students, their part-time status (if not their GPA) more often

makes them ineligible for first-class standing and scholarly awards. Understandably, part-time attendance also delays graduation. The requirement to withdraw for academic reasons is our final indicator of academic performance.

Annual decisions concerning withdrawal (and promotion) are made every spring at the completion of winter session. These decisions are based on a minimum of nine course weights (three term courses) taken during winter session. They require the calculation of a quality index (QI). The QI is calculated by dividing the sum of course weights passed by the sum of course weights attempted and multiplying the result by the sessional GPA. The QI, like the GPA, ranges from 1 to 9. A QI of 2.0 or less in a single winter session or a QI between 2 and 4 in any two winter sessions results in a requirement to withdraw. In spite of poor performance, part-time students completing only nine course weights in winter session escape compulsory withdrawal and avoid having the notation "required to withdraw" permanently imprinted on their academic transcript.

The primary independent variable is admission status: admission on the basis of high school matriculation (coded as 0) or NMA standing (coded as 1). Our primary objective is to investigate differences between matriculants and NMAs in their performance at university. In a sense we want to investigate whether the maturity (older age) of NMAs is in some sense a viable substitute for matriculation.

Age is also an important variable. It is measured as age in years on admission. For September admission, students must be 21 years of age before September 30, and for January admission (of which there were few), the 21st birthday must occur before January 30. However, age is confounded by admission status, because age is one of the principle criteria for admission as an NMA. All NMAs were 21 years of age or older, and all but 25 of the 579 matriculant admissions were under 21. Still, if maturity has a pervasive effect on performance, age differences may have a positive effect on GPA in admission cohorts. A year or two can mean a great deal to the matriculant admissions, and NMA admissions in their late 20s and beyond may be considerably more mature than those in their early 20s. We have to be aware of the close relation between age and admission status when interpreting our results.

Past academic performance is also obviously relevant to performance at university. However, we were limited in our selection of indicators, because NMAs by definition lacked relevant high school records that were comparable to matriculation admissions. Besides age (and a lack of recent academic failure), the only other criterion for admission of NMAs in 1985-1986 was grade 12 English (English 30). It was therefore used to examine the effects of the only indicator of prior performance used in the admission decisions. Matriculation admissions were scored on the basis of the percentage grade they received in high school English as reported on secondary school transcripts. Nonmatriculant admissions were also given their percentage grade in high school English if they had taken the course. The vast majority of NMAs (approximately 90%) satisfied the English criterion with a high school English. About 9% of the NMAs satisfied the English admission criterion by completing the Writing Competence Test. Grades on this test were converted to percentages using the formula devised by the Registrar's Office.[3] Only four students com-

pleted a university level English course. If NMAs satisfied the admission English criterion by completing a first-year university level English course, their grades on the 9-point scale were reconverted to percentages using the University's conversion formula.[4]

We also were able to add a variable representing whether or not the student had a deficiency in the requirement for a language other than English (LOE) on admission (a deficiency is coded as 1; no deficiency is coded as 0). This information was recorded as part of the admission ruling for this and earlier cohorts because all Arts students, as a condition of graduation, were required to present credit in a language other than English either at the senior matriculation or university level. Since 1990, the LOE has been introduced as an admission requirement at the senior matriculation level and as a degree program requirement at the university level. LOE deficiency, therefore, gives us a second crude indicator of prior performance that has additional policy relevance. Presumably, the assumption is that if a mature student has an LOE credit, they are more likely to perform better in the Faculty of Arts.

Results

Admission Status and Overall GPA

Our analysis focuses on the 579 matriculants and 332 NMAs admitted to the Faculty of Arts during the 1985-1986 academic year.[5] Contrary to our predictions, the 332 NMAs did not perform better at university than the matriculants (Table 1). The NMAs' overall GPA was 1.06 lower than that of the matriculants. The simple regression of GPA at the end of four years on type of admission (with admission status coded NMA=1, Matric=0) reconstructs these same results a little differently. The intercept was 5.97 (the mean for the Matrics) and the slope was –1.06 (the mean for the NMAs was 1.06 lower). This simple admission status difference accounted for 7.6% of the variance in GPA.

Table 1
Student Characteristics by Type of Admission

Student Characteristics	*Total Sample*	*High School Sample*	*Mature Admission (NMA)*
Average Age (Yrs.)	22.2	18.3	27.4
		Percent	
Female	60	61	57
English 30 grade	71	74	65
LOE deficiency	49	28	84
Won Award	5.4	7.1	2.4
Graduated	33	40	19
First-class Standing	16	21	8
Ever forced to withdraw	12	12	13
Overall GPA	5.59	5.97	4.91
Std. Dev.	(1.87)	(1.52)	(2.21)
Sample Size	911	579	332

Other indicators of overall performance are consistent with these results. This is expected because the requirement to withdraw as well as the achievement of awards, first-class standing, and graduation are partly dependent on GPA. By the end of four years, matriculants were more likely to have won awards (7.1% vs. 2.4%), to have achieved first-class standing in any year (21% vs. 8%), and to have graduated (40% vs. 19%). High school matriculants were only slightly less likely to have ever been forced to withdraw (12% vs. 13%).[6] On the average, matriculants had only slightly better English 30 marks than NMAs (74% vs. 65%), although as we would expect, they were much less likely to have been admitted with an LOE deficiency (28% vs. 84%).

The Effects of Age and Sex

When we add the demographic factors of age and sex to the regression (Table 2) we increase the R-square to .110. For the whole entering cohort, age was virtually uncorrelated with GPA (Table 2, r= –.065). Older students did not have higher GPAs. However, when we controlled for admission status we did observe the predicted positive effects of age on GPA (St. Beta=.217). Admission status suppressed the positive effects of age on GPA. Controlling for gender and admission status, for each year increase in age, predicted GPA increased by b= .063. For a mature student entering in his or her early 30s instead of early 20s (10 years difference in "maturity"), we would predict a GPA about .63 higher.

The observed zero-order effects of gender on GPA were small and negative (r= –.109). This weak tendency for males to have somewhat lower GPAs than females was not greatly altered with the addition of controls for age and admission status (St. Beta=–.087). GPA for females averaged only slightly higher than males (about .331 grade points), controlling for admission status and age. This confirmed our expectation that there would be only small gender differences in GPA. However, when age and gender were controlled, the differences between matriculant and NMA admissions increased to 1.626 from the observed difference of 1.060 and the zero-order correlation (r= –.275) increased to a standardized beta of –.419. This suppressor effect was produced largely by the strong positive correlation (r=.679) between age and admission status. As we expected, age and admission status were partially confounded,

Table 2
Overall Grade Point Average: Admission Status, Age, and Sex

Variable	*b*	*St. Error*	*St. Beta*	*Tolerance*	*r*
Regression Results					
Admission Status (NMA=1, Matric=0)	–1.626	.166	–.419	.536	–.275
Age (years)	.063	.012	.217	.536	–.065
Sex (Male=1, Female=0)	–.331	.120	–.087	.994	–.109
Constant	4.939	.248			

R^2 = .110 ; *df* = 3, 906.

because age was used as an admission criterion for NMAs. However, despite the strong correlation between age and admission status, there was no evidence of substantial collinearity. Tolerance remained above .5 for both variables (Table 2), and standard errors for the partial correlation coefficients were not inflated. Further tests revealed no significant interaction effects among age, gender, and admission status as they affected the overall GPA.[7]

Preuniversity Performance and GPA

To see if preuniversity academic performance accounted for the performance difference of Matrics and NMAs we added English 30 grades and LOE deficiency to the regression equation (Table 3). We expected both to have significant effects on GPA, and we also expected them to account for part of the effects of admission status.

Adding indicators of prior performance increased the R-square to .184. When we controlled for the two indicators of previous academic performance, the effect of type of admission was reduced to .892 points. Prior academic performance explained a little less than one half of the admission status differences observed in Table 2. Still, admission status had substantial effects on GPA (St. Beta=–.230). Gender effects now had virtually no effect on GPA (St. Beta=–.032). The effect of age remained relatively unchanged (b=.060; St. Beta=.205). LOE deficiency on admission had little effect on GPA (St. Beta=–.090). Those without a language other than English had average GPAs that were .337 lower. English 30 grades had a larger partial effect (b=.056; St. Beta=.288). A difference of 10 points on a student's English 30 grade would predict a GPA that was .56 points higher.

First-year Performance and GPA

Finally, in order to examine the effects of early performance at university on later GPA, we recomputed our dependent variable to represent average GPA for courses taken during years two through four. We then included the average of grades obtained in the students' first year (1985-1986) as a predictor. As we would expect, the R-square for the equation increased considerably (to .400).

Table 3
Admission Status and Overall GPA

Variable	*b*	*St. Error*	*St. Beta*	*r*
Regression Results				
Admission Status (NMA=1, Matric=0)	–.892	.185	–.230	–.275
Age (years)	.060	.012	.205	–.065
Sex (Male=1, Female=0)	–.122	.119	–.032	–.109
English 30 Grade	.056	.007	.288	.374
LOE Deficiency	–.337	.138	–.090	–.252
Constant	.874	.554		

R^2 = .184 *df* = 5, 904.

First-year grades were strongly correlated with GPA for years two through four (r=.602), and first-year GPA was an excellent predictor of later GPA (St. Beta=.576; b=.938).[8] When first-year grades were controlled, the effect of admission status was reduced somewhat (b=–.711). Still, the effect of admission status persisted throughout years two through four even when we controlled for initial performance.

When first-year performance was controlled, English 30 grades had little effect on subsequent university grades (St. Beta=–.032). LOE deficiency had a somewhat larger effect (St. Beta=–.098), but this effect was still small. These results suggest that any effects of these two factors was largely captured by first-year GPA. The effect of first-year grades also appeared to account for most of the observed effect of age on later GPA (St. Beta=.032). Controlling for first-year GPA (and the other variables in Table 4), the partial regression coefficient for age was reduced to almost zero (b=.008, compared with b=.060 in Table 3). The greater maturity of older students helps them initially, but it does not appear to help them more as they progress through their university studies.[9]

Discussion

Contrary to our expectations and those of the University of Alberta, the NMAs from this cohort performed substantially more poorly than Matriculation admissions. At the end of four years, the less qualified mature students averaged a full grade point less than Matric admissions. They still scored an average of .89 grade points less when overall GPA was adjusted for demographic and prior performance factors.

Gender differences were negligible. By contrast, age differences in performance at university were substantial. Yet age had little effect on later performance in years 2 through 4. It therefore appeared that the principal effect of age differences occurred early in a student's university career. English 30 grades had similar effects. They contributed to the prediction of overall GPA, but when GPA for years two through four was regressed on first-year GPA,

Table 4
Determinants of GPA for Years 2-4

Variable	*b*	*St. Error*	*St. Beta*	*r*
Regression Results				
First Year GPA	.938	.04	.576	.602
Admission Status (NMA=1, Matric=0)	–.711	.237	–.125	–.278
Age (years)	.008	.015	.032	–.066
Sex (Male=1, Female=0)	.117	.009	–.032	.267
English 30 Grade	–.009	.009	–.032	.267
LOE Deficiency	–.533	.177	–.098	–.267
Constant	.929	.708		

R^2 = .400 df = 6, 868.

English 30 grades had little impact. The effect of English 30 appeared to be "captured" by first-year performance. Whether or not the student had a deficiency in a language other than English (LOE) had little effect on GPA.

What implications do these results have for formulating policy with respect to the admission of NMAs? The consistently lower averages of NMA students compared with regular Matric admissions are persuasive. Whatever maturity is associated with NMA status apparently does not make up for lack of qualifications. However, age did have positive effects on initial performance (first-year GPA) and NMA students are generally older than Matric admissions. In order to make up for this observed GPA deficit, NMAs would have to be about thirty years old when admitted. From this perspective, the age criterion for NMA admission would have to be increased substantially.

In 1990 the Faculty of Arts did raise the minimum age for NMA admission from 21 to 24 years. This was part of a more restrictive move to manage enrollments while maintaining academic standards (Committee on Admissions and Transfer, 1988). A closer examination of the 1985-1986 study cohort, by age groups and overall GPA, is instructive (Table 5). For the Matric students, the youngest admissions (16-17 years) have a higher overall GPA than any of the age groups except for those few students who were 24 years or over on admission. This apparently curvilinear effect of age on GPA for Matric admissions may have been something of an historical anomaly. Cohorts entering at age 18 through 23 contain a substantial proportion of the increasing numbers of students who during the 1980s took more than one year to complete grade 12 (Alberta Advanced Education, 1987). This includes students who went back to get particular subjects, but it also includes a substantial proportion of students who retook courses in order to raise their average. The difficulties that these students experienced in high school may simply continue with poorer performance in university. It is not until matric students reach the age of 24 years that their performance exceeds that of students who completed high school in the standard minimum time (16- to 17-year-olds).

Table 5
Age and Overall GPA by Admission Status

Age (years)	*Matric*	*N*	*Nonmatric*	*N*
Overall GPA				
16-17	6.22	130	*	
18	5.94	344	*	
19	5.67	61	*	
20-23	5.45	28	4.33	127
24-29	6.59	9	4.99	111
30-39	7.03	6	5.37	68
40 and over	6.00	1	6.22	26
Total	5.97	579	4.91	332

*All NMAs were 21 years of age or older.

Table 6
The Effects of Age on Overall GPA: One-way ANOVA for Matric and Nonmatric Admissions

	Admission Status	
	Matric	*Nonmatric*
F	2.32	7.51
Sig. F	.032	.000
Eta^2	.024	.064

For the nonmatric admissions, only those who are 40 years of age or older had higher overall GPAs than any of the matric admission age groups. The youngest nonmatric admissions, who constituted over one third of the nonmatric cohort, had by far the poorest performance. Based on our results, we would expect the change in prescribed age for NMA admission to improve GPA although somewhat less than was hoped for. It appears to take a good deal of maturity and experience, at least as reflected in chronological age, to compensate for lack of academic preparation. Still, age did improve performance for both matric and particularly nonmatric admissions (Table 6).

The Faculty of Arts introduced another change in 1990. The language other than English was prescribed as an entrance requirement for NMAs. Fulfillment of this requirement preadmission would obviously facilitate the completion of the Arts degree language requirement. However, this additional entrance requirement (in combination with a higher minimum age) had the effect of reducing NMA admissions to a comparative trickle (51 in 1990-1991, 55 in 1991-1992, 49 in 1992-1993). Based on our results, we would not expect this additional entrance requirement to improve subsequent university performance in other subjects. LOE deficiency had little effect on GPA.

Why did the Faculty of Arts at the University of Alberta not get the academic performance of nonmatriculated admissions that previous studies would have predicted? We can only speculate, but the answer seems to lie in the reflexive experience that screening procedures make available. The successful NMA programs in Canada (Perkins, 1971; Wilson & Lapinski, 1978), Australia (Barrett & Powell, 1980; Eaton & West, 1980) and Britain (Smithers & Griffin, 1986; Walker, 1975) all employed rigorous methods of personalized screening as part of the admission process. It may not matter if the entrance test (whether this is a critical essay or an exam taken after a compulsory lecture on an unknown topic) is even scored (Barrett & Powell, 1980). What matters more is that the method used should acquaint "mature" applicants with the nature and demands of university study and supply them with the chance to assess their own abilities, interests, and motivation.

A higher participation rate in university education is one of the most important achievements in the past three decades. Bureaucratization and standardized decision making accompanied the shift to mass education in postsecondary institutions. The University of Alberta was not alone in abandoning labor-intensive individualized screening of NMA admissions in favor of minimalist formal criteria. The lesson may be that access and academic

performance have to be carefully balanced. Increased access for mature students cannot be successfully implemented without the application of significant organizational resources.

Notes

1. Besides "regular" NMAs, the Faculty of Arts also admits applicants without matriculation who (a) were previously admitted to another university on the basis of mature standing or (b) had successfully completed the first year of a technical or vocational program in Alberta. These two admission categories include only a relatively small number of freshman students (51 and 67 respectively in 1985-1986) and they are not necessarily "mature" (21 years or older). Both categories are omitted from the study.
2. Cost considerations precluded collecting supplementary information directly from students.
3. There were only two passing grades: S (satisfactory) was scored as 85%; MS (mostly satisfactory) was scored as 65%. If students received an MU (marginally unsatisfactory) or U (unsatisfactory), they would not satisfy the admission English criterion.
4. The conversion formula assigns each stanine score the midpoint of the percentage range to which it refers: 9=95, 8=85, 7=76, 6=69, 5=62, 4=54. Grades below 4 would not satisfy the admission criterion.
5. Interpretation of results emphasizes the magnitude of effects, because our sample was not designed to be directly representative of a larger population. Tests of significance are reported, but they are less important than the magnitude of the effects observed (Cohen, 1991).
6. Closer examination of the record reveals that a large part of the reason for this particular finding is that nonmatriculants are more likely to withdraw voluntarily. An indication of this is that more NMAs (17% vs. .04%) actually completed fewer than three courses in four years. All but five of the students falling in this category failed to accumulate any grade points after the first year.
7. Given the policy relevance of age 21 (admission criterion for NMA in this cohort) and 24 (the new 1990 criterion), we examined separate regressions of sex and admission status on GPA for both of these age splits. No significant differences in effects were observed.
8. We might note that there is no evidence of grade inflation in these results. Had there been substantial grade inflation from the first year to subsequent years, we would have expected to observe a partial slope for first-year GPA that was greater than 1.0.
9. There was no significant interaction between admission status and the effects of initial performance on later performance.

Acknowledgments

This research was supported in part by Student Temporary Employment Program Grant LPT 216576. The authors gratefully acknowledge the assistance of Vic Stobee in the Faculty of Arts and Marj McMullen in the Office of the Registrar. We also thank an anonymous reviewer for helpful comments.

References

Alberta Advanced Education. (1987). *Enrolment and population trends: A background paper.* Edmonton, AB: Research and Information Support.

Alberta Advanced Education. (1990). *Responding to existing and emerging demands for university education: A policy framework.* Edmonton, AB: Author.

Anisef, P., Okihiro, N., & James, C. (1982). *Winners and losers.* Toronto, ON: Butterworth.

Barrett, E., & Powell, J.P. (1980). Mature age unmatriculated students and the justification of a more liberal admissions policy. *Higher Education, 9,* 365-383.

Cohen, J. (1991). Things I have learned (so far). *American Psychologist, 45,* 1304-1312.

Committee on Admissions and Transfer. (1988). *Report of the subcommittee on non-matriculated admissions.* Edmonton, AB: University of Alberta.

Darling, A.L. (1985). The admission of non-matriculated adults to university credit courses. *Canadian Journal of Higher Education, 15*(2), 91-97.

Eaton, E., & West, L. (1980). Academic performance of mature age students: Recent research in Australia. In T. Hore & L. West (Eds.), *Mature age students in Australian higher education* (pp. 48-68). Victoria, Australia: Monash University.

Jackson, L.S., Small, J.M., & Zelmer, A.E. (1985). Mature student provisions at the University of Alberta. *Canadian Journal of University Continuing Education, 11*(2), 64-74.

Johnson, M.L.,& Walberg, H.J. (1989) Factors influencing grade point averages at a community college. *Community College Review, 16*(4), 51-60.

The Mature Student Entry Advisory Committee. (1983). *Final report*. Burnaby, BC: Office of the Registrar, Simon Fraser University.

McLeish, J. (1962). The age factor in adult learning. *International Review of Adult and Youth Education, 14*, 26-45.

Perkins, S. (1971). Success of mature age students in university studies. *Alberta Journal of Educational Research, 17*, 105-112.

Smithers, A., & Griffin, A. (1986). Mature students at university: Entry, experience and outcomes. *Studies in Higher Education, 11*(3), 257-268.

Vaselenak, M. (1970). Admission of mature non-matriculated students into a degree program. *McGill Journal of Education, 5*(1), 85-96.

Walker, P. (1975). The university performance of mature students. *Research in Education, 14*, 1-13.

Wilson, L., & Lapinski, B. (1978). A longitudinal study of mature students in credit programs. *Canadian Journal of Continuing Education, 4*(2), 22-27.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 435-461*

June Beynon

and

Kelleen Toohey
Simon Fraser University

Access and Aspirations: Careers in Teaching as seen by Canadian University Students of Chinese and Punjabi-Sikh Ancestry

This study, based on interviews with 34 Canadian university students, 22 of Chinese ancestry and 12 of Punjabi-Sikh ancestry, revealed that cultural influences (i.e., parents' views about the desirability of teaching) and structural barriers (i.e., discrimination in schools and insufficient instructional support in academic English) are implicated in individuals' perceptions of teaching careers. Increasing participation in teaching requires a theoretical perspective that includes these factors as well as the views of individuals who interpret and react to them. This research shows there is diversity in the way these factors operate in the two ethnic groups, as well as diversity based on gender and individual perspectives and experiences. Recommendations made for increasing participation in teaching attend to this variability.

Cette étude basée sur des interviews accordées à 34 étudiant(e)s universitaires canadien(ne)s dont 22 d'ascendance chinoise et 12 d'ascendance punjabi-sikh, révèle que les influences culturelles (ex., l'opinion des parents concernant l'attrait à l'enseignement et les barrières structurelles (ex., la discrimination dans les écoles et l'insuffisance du support dans l'enseignement de la langue anglaise) sont tenues en ligne de compte dans leur perception de l'enseignement comme choix de carrière personnel. La participation croissante dans l'enseignement requiert une perspective théorique qui doit inclure ces facteurs ainsi que les opinions des individus qui les interprètent et qui y réagissent. Cette recherche démontre qu'il y a une diversité dans la façon qu'opèrent ces facteurs dans ces deux différents groupes ethniques ainsi qu'une diversité fondée sur le sexe, et les points de vue et les perspectives et les expériences individuels. On recommande d'accroître la participation dans l'enseignement pour répondre à cette variabilité.

Introduction

> Having minority teachers would definitely help minority students in adapting to Canadian culture. Minority teachers would be more aware of the problems students might be having and how to help them. (Peter, Canadian university student of Chinese ancestry)

June Beynon is an associate professor in the Faculty of Education where she teaches courses in multicultural and antiracist education. She has also worked as the multicultural education officer for the Vancouver School Board. She is currently researching First Nations teachers' perceptions of their roles as social change agents.
Kelleen Toohey is an associate professor in the Faculty of Education where she teaches ESL methodology. Currently she is conducting research in the integration of second language learners into Anglophone classrooms.

To respond to the needs of minority students, as Peter has outlined, and to address concerns about equity of access and equity of representation in employment are the prime reasons that educational practitioners and policy makers are currently advocating recruitment and training of minority teachers.[1] In addition, having teachers who are representative of the social, cultural, and linguistic diversity of Canadian society provides the possibility for all students to learn from individuals who represent culturally diverse perspectives. In our role as teacher educators, we feel obliged to discover the character of barriers at the university that could prevent minority group individuals from entering teacher education programs. Students' perceptions of the desirability of teaching as a career, as well as their views of the particular factors that enhance or inhibit access to this career, may be useful to us and potentially other institutions in understanding the reasons for current underrepresentation and what we might do to address the problems.

The purpose of this study was to explore the perceptions and articulations of experiences affecting career choices made by Canadian university students of Chinese and Punjabi ancestry (the two most populous visible minority communities in British Columbia). Previous research corroborated our impression that Canadians of Chinese ancestry are not well represented in our teacher education program (Beynon, Toohey, & Kishor, 1993). We found that although Canadians of Chinese ancestry were well represented in undergraduate enrollments in proportion to their provincial population figures, applications and acceptance to the Simon Fraser University teacher education program of Canadians of Chinese ancestry showed underrepresentation. Current representation in our teacher education program of Canadians of Punjabi-Sikh ancestry was also lower than their university enrollment, which itself was proportional to provincial population figures.[2] Based on these demographic data, we began an inquiry into the factors that might explain the underrepresentation of these two groups of students.

Our previous research suggested that family members' expectations, parents' wishes, and ethnic group expectations were significant influences on the career choices of visible minority students in British Columbia and were more significant for them than for Canadians of Anglo-European ancestry.[3] The questionnaire research showed us that the important factors in choosing a career related primarily to personal, familial, or "cultural" matters and not to externally imposed barriers to career selection. Although students made occasional reference in the questionnaires to issues of discrimination or institutional barriers to teacher education, this format did not appear to elicit many of these sorts of disclosures. The present study was designed to find out more about how a group of Canadian university students of Punjabi and Chinese ancestry saw these factors.

Background

Canadian social policy (Abella, 1984) commonly posits similarities between visible minorities, women, individuals with disabilities, and First Nations as groups in Canadian society who have experienced discrimination by employers and possibly educational institutions. The report suggests that members of these groups should be targeted for training and employment. Social theorists Carty and Brand (1993), McCarthy (1988, 1990) and Ogbu

(1991), on the other hand, advocate the need to look carefully at the specific circumstances of groups whose members experience discrimination. If discrimination is experienced differently by different groups (or by individuals within groups), then efforts to disrupt discrimination may well need to be various. If the reasons that students are not entering teaching careers vary from one ethnic group to another, for example, it would be important to find out what their different reasons are.

McCarthy (1990) makes the point that liberal discourse about race and education overrelies on "culture" and values when explaining inequity, thereby, "abandoning the crucial issues of structural inequality and differential power relations" (p. 56). At the same time, Neo-Marxist approaches, he argues, overrely on economic explanations and imply that individuals make few if any choices or have little agency. A purely cultural view would give the responsibility for change largely to the individual and cultural group. A purely structural view would see the responsibility for change placed largely on economic or other institutions. For example, sociological research on relationships between social class and career choice indicate that "educational plans and vocational decisions are largely determined by socio-economic background" (Li, 1988b, p. 75).

For more than two decades paradigms for multicultural and antiracist education in Canada and the United States have designated the school as an institution with a central role in dismantling inequity. This paradigm casts individual parents, teachers, and students as active and empowered in bringing about change (Banks, 1991; Cummins, 1986; Pang, 1991; Ruiz, 1991; Sleeter, 1991). Specific agendas, theoretical and practical, for transformation of teacher education are also fundamental to restructuring schools, and in turn society at large (Baptiste, Baptiste, & Gollnick, 1980; Bennett, 1988; Beynon & Warsh, 1993; Gollnick, Osayande, & Levy, 1980; Grant, 1983; Lynch, 1986; Martin, 1991).

McCarthy (1988) would have us consider both the personally articulated, "the symbolic, signifying, and language dimensions of social interactions" (p. 246), and the structural aspects, as one considers action around amelioration of racial inequality. The emphasis on personal articulations is consistent with the interactionist approach to social research that focuses on the (diverse) descriptions, explanations, and meanings given by the individual participants. In this view, "to understand people's behavior we must use an approach that gives us access to the meanings that guide their behavior" (Le Compte & Preissle, 1993, pp. 31-32). As Hammersley and Atkinson (1991, p. 7) point out, "It is these interpretations, continually under revision as events unfold [that] shape their actions." Accordingly, in this study, even though we start out knowing, for example, that parental influence is salient, we do not assume that the same parental expectations will have the same effects on all students, even if they are of the same gender and ethnic background.

Our research considers how members of the two most populous visible minority groups in British Columbia describe their decisions about their occupational futures with special attention to teaching. The views they have of themselves with regard to teaching and the influence they perceive their ethnic group membership to have are investigated. Do these individuals feel that they

have encountered discriminatory barriers to their pursuit of education and career? Are some members of the ethnic group more constrained than others in their occupational choices? In what ways do answers to the above questions reveal patterns by ethnic group, by gender, by class, or by some other grouping, and in what ways are responses unique to the individual? Finally, how might this information help us to apply principles of equity in teacher education? By looking at students' views we hope to offer new grounds for dialogue and action in the debate between cultural versus structural analysis.

We have examined in detail the transcripts of interviews of 22 Canadian students of Chinese ancestry and 12 Canadian students of Punjabi-Sikh ancestry. In drawing our data from individual interviews we are aware of critiques of this longstanding research tool that caution about the dangers of generalizing information from individual interviews and transforming it into assertions about classes or groups of individuals in society (Clifford, 1988; Van Maanen, 1988). We are reminded as well that interviewing involves the subjectivities and agencies of interviewer as well as interviewee and that the "answers" of an interview are highly complex and partial (Hale, 1991; Stacey, 1991). Recognizing this complexity and seeing this research primarily as a way to speak to ourselves about how the university might address problems that it is only gradually coming to recognize, we have limited our conclusions, seen them as tentative, and make few recommendations for others. By working with relatively small numbers of participants in lengthy interviews, we also believe it has been possible to be attentive both to expressions of difference and to common themes.

Clearly Canadians of Chinese ancestry have come from many different places at many different times in the past, and the personal, familial, and "ethnic" group experiences of individuals classified as "Chinese Canadian" vary widely.[4] Differences between the experiences of this diverse group and those of Canadians of Punjabi-Sikh ancestry, of course, are wide as well.[5] Taking into account all these differences is clearly impossible within the constraints of the present report. Rather, this study focuses on perceptions of teaching as a career by some members of these groups.

Methods and Procedures

The interviews that form the data base for this research were conducted with individuals who indicated their willingness to volunteer on a questionnaire completed by a total of 751 university students at the University of British Columbia and Simon Fraser University (Beynon, Toohey, & Kishor, 1993). The interview questions were designed as follow-up to the topics raised in the questionnaire, allowing opportunities for the students to respond more fully and to indicate the nature of the connections, if any, among their responses regarding family influence, institutional barriers, and career choices.[6]

Interviews were 45 to 60 minutes long. The interviewers were four Canadian women of South Asian, Native Hawaiian, and Anglo-European ancestry. Their experience as interviewers ranged from one to 16 years. After interviews were completed, interviewers discussed the work. Minority interviewers expressed less (retrospective) discomfort in interviewing minority students about discrimination than did interviewers of Anglo-European background. Similarly, they felt visible minority students displayed little unease in speaking about

discrimination with minority interviewers. Interviewers of Anglo-European ancestry were more uneasy asking, and they felt students were more uneasy telling them about discriminatory events. Nevertheless, minority students did report discriminatory events to nonminority interviewers.

Tapes were summarized and transcribed by four women of Anglo-European ancestry (different from the interviewers) with considerable experience in psychology, education, writing, and communications. Each transcriber worked with material from each interviewer and each designated group. They used the format of the interview as their guide to produce a four- to five-page overview with selected verbatim quotations to introduce key points. After reading the transcripts, the authors (two women of Anglo-European ancestry) returned to the tapes to make additional verbatim transcriptions. As a final step in verifying the accuracy and acceptability of the transcripts we sent each to the respective interviewee and asked him or her to correct any inaccuracies. Modifications were minor, as they tended to embellish rather than change the points made. In one instance a respondent indicated that the transcript did not accurately reflect her meanings, and that respondent's interview was not used in this analysis.

We grouped the data first by ethnicity and then by gender. Two comments by students of Chinese ancestry indicate the general awareness that gender was important in career selection:

> Traditionally if a girl ends up being a teacher it's great; it's considered one of the best career choices for them, but if a boy ends up teaching that is not so good. A female teacher is prestigious but for a male it is not. (Joshua)

> My mom doesn't really care whether I go into anything as long as I get a degree from university. My dad, he actually said, just go get married, who cares about education ... you'll eventually end up getting married ... I have three brothers who are quite good at school so they [her parents] don't really care about me. (Sandra)

Burt, of Punjabi-Sikh ancestry, comments on differences between expectations for boys and girls in his community:

> Things are not equal between the sexes in my cultural group and many parents in the Sikh community do not want their daughters to attend university until they are married—if at all.

The sorting by gender within ethnic group is supported by a growing feminist literature that illustrates how minority women's experiences in family, education, and career development diverge in important ways from the experiences of their male counterparts (Bannerji, 1993; Carty & Brand, 1993; Ng, 1993). The report of the students' interviews below makes apparent, we believe, differences in the perceptions, experiences, and aspirations of women and men in these groups.

Our objective in reviewing the interview data was to ascertain both common and diverse themes that emerged from the individual narratives. Although we structure the themes here in skeletal form, nuances emerge in the students' narratives around these issues. The selection of quotations illustrates the salience of the patterns as well as the uniqueness of the individuals whose

narratives revealed these patterns to us. All names have been changed to maintain confidentiality.

We have a great deal of data and finding ways to report these data respecting the particularity of individual's accounts but also considering the length constraints of an academic journal article has been challenging. We make brief general comments about aspects of the information received from students that is relatively straightforward and not necessarily illuminated by the "very words" of the respondents. The body of the data report contains excerpts from the interviews, some of which are particularly apt illustrations of general points and some of which eloquently convey unique perspectives.

Data Analysis

In analyzing our interview data for themes, three factors appeared most salient: although parental influence is one major reason that these students do not choose teaching as a career, we also found that lack of confidence in their own English language skills and experiences of prejudice in the schools are also key influences. A key issue for most Canadian students of Chinese ancestry is parental influence in career choice. Several also identified lack of English proficiency as a barrier to pursuing teaching as a career. Few mentioned racism and discrimination. In contrast, among Canadian students of Punjabi-Sikh ancestry, parental influence and language issues are not identified as impediments to choosing teaching as a career, but racism and discrimination in the schools are. None of the students mentioned that they perceived discrimination in teacher education programs, but this issue was discussed in our earlier publication.

There is no well-developed body of Canadian literature on the area we are studying, but the themes that emerged in our data echo themes in a clearly focused United States body of literature on school achievement for Americans of Chinese and Punjabi ancestry. These studies (as does our work) document that students of Chinese ancestry enter teacher education less frequently than students from other ethnic groups (Peng, 1985; Vetter & Bapco, 1984). The common themes that emerge are the powerful influence of parental expectations on school achievement and career selection, the relationship of language skills to school achievement and career selection, and the reality of prejudice in the schools (Chinn & Wong, 1992; Hsia, 1988; Pang, 1990; Vetter & Bapco, 1984). These themes are also relevant for students of Punjabi ancestry (Gibson and Bhachu, 1991; Singh Ghuman, 1993). A brief discussion of that literature follows the report on our data.

General Comments

Age

The participants as a whole were relatively young. The ages of the men of Chinese ancestry ranged from 19 to 25. Women's ages ranged from 19 to 23. In the Punjabi-Sikh case, ages for four men ranged from 20 to 22; one was 33. The women's ages were from 20 to 22.

Class

The class background (defined here by parental occupation) of the students ranged from laborer to professional.[7] In the case of the men and women of Chinese ancestry, parental occupations included cook, baker, seamstress,

homemaker, accountant, teacher, pharmacist, doctor, and entrepreneur. In the case of the men and women of Punjabi-Sikh ancestry parental occupations included millworker, farm laborer, machinist, homemaker, bus driver, teacher, social worker, and engineer.

Financial inaccessibility of postsecondary education is cited often in sociological research as a critical factor preventing minority youth from attaining high socioeconomic status through the pursuit of managerial or professional careers (Breton, 1970; Buttrick, 1977; Gilbert & McRoberts, 1977; Li, 1988b). Quantitative data from our previous study indicated that this is not necessarily the case for students from the two groups studied, who come from a wide range of socioeconomic backgrounds and are well represented in university (Beynon et al., 1993).[8] Class background of the students may well have an effect on both the sorts of careers they select and the reasons they find compelling for choosing those careers. We report parental occupations in the sections on parental influence, but the data from this study did not reveal any clear relationships between parental occupation and perceptions of the desirability of teaching as a career.[9]

Countries of Origin

Related to issues of class are the diverse economic histories of families in their countries of origin, which may differentially influence career choice. In contrast to earlier policies emphasizing recruitment of labor, Canadian immigration, particularly since 1979, has emphasized recruitment of business immigrants with a large amount of capital to invest in Canada (Li, 1988a, p. 125). For Chinese immigration this has meant an increase in business immigrants from Hong Kong. However, of the students we interviewed there were only two instances (of students of Chinese ancestry) where parents might be identified as post-1979 business immigrants.[10]

The career choices the students had made are reported in Table 1.

Canadian Women of Chinese Ancestry

Only one of the 12 women in this sample was planning on teaching as a career. We begin with a profile of her experiences. This is followed by an analysis of the other 11 women who framed their reasons primarily in terms of parental influence and English language proficiency. In this group of women only two reported experiences of prejudice and discrimination in their own schooling.

Table 1
Career Choices of University Students
of Chinese (cc) and Punjabi (cp) Ancestry

	Law	*Business*	*Science*	*Health*	*Engineering*	*Public Relations*	*Arts*	*Education*	*Police Work*	*n*
cc (female)		4		4			3	1		12
cc (male)		4	3	1	2					10
cp (female)	1	2	1			1	1		1	7
cp (male)	1	2	1					1		5

Lydia, who is enrolled in a teacher education program, is 22 years old and immigrated as an infant from Hong Kong. She told us about when she first thought about becoming a teacher.

> In grade 1 my teacher was reading us a story and I liked what she was doing so much from that moment all I thought about was teaching.

In spite of difficulties encountered in school and ambiguities in her parents' expectations she was still committed to pursuing a career in teaching.

> In high school it was very difficult for me and a lot of teachers told me that I wouldn't get very far, that I wouldn't get past high school. I resented that. My grade 9 and grade 10 math and French teachers told me not to go on in school. If I teach I want to be more effective; I want to be more nurturing. I want to let students know that if they aren't doing well right now it's OK, there's always a second chance. You can always improve.
> [My parents] didn't think I could do well enough at university, although they think teaching is an honorable job and are fairly satisfied with my choice [to be a teacher], but they think I am going to be beaten up [working in a school]. (Lydia, parents' occupations: grocers)

Parental Influence

The women cited strong and pervasive parental influence and expectations, and in general parents seemed more willing to accept teaching as a career choice for females than for males. They distinguished two realms in which parental influence and expectations operated: university attendance and career selection. Ten of the 12 women's narratives begin with strong statements that as long as they attended university, their parents "didn't mind" what career they followed. But as their narratives unfolded it became apparent that their parents favored careers in business that were deemed practical and financially secure. Emphasis on occupations deemed prestigious did not appear to be as pronounced as it was in the case of the men. However, there were comments indicating that for women university attendance in and of itself was considered prestigious and desirable with 10 of the 12 women citing explicit parental expectations about university attendance. Typical was:

> Ever since I was a kid it was assumed I would go to University. The Singapore mentality on the whole is, your job is to study, to get good grades and that is that. (Lauren, father: entrepreneur; mother: homemaker)

> I just knew that I had to go to University; it wasn't a question of whether I didn't or not. It wasn't anything I could pick. I couldn't say that I wanted to take a year off and go to work. It was assumed I would go to University. (Alice, father: pharmacist; mother: homemaker)

> My parents would rather have me get a degree than just a diploma or a certificate. When they talk to their friends and relatives and say "Oh, my daughter goes through university" instead of just a college or an institute, it just sounds better. (Sandra, father: retired; mother: homemaker)

Eight of the 12 women noted parental approval (in varying degrees) of teaching as a career for them, occasionally referring to someone in their family who already was a teacher. A sample of their comments:

> They think it's a good occupation and they wouldn't mind at all. (Corinne, father: gasfitter; mother: property manager)

> My father gave me his blessing with teaching because my mom is a teacher. (Lauren, father: entrepreneur; mother: homemaker)

> My parents would approve of teaching, just as equally as nursing. My sister is a teacher. (Lucia, father: chef; mother: worker)

> My parents would be neutral. It is a secure profession. To my mother's generation teaching was the most natural thing to do for a girl. I'm not too sure about my dad, but my mom wouldn't mind. (Carol, father: translator; mother: office worker)

Two of the 12 women reported their parents would not approve:

> My family would approve of teaching only if there is a high demand for teachers in the marketplace. It is not a preferred career and the first thing they examined was the financial aspect. (Laura, father and mother both retired)

> My parents would think there was nowhere to go as a teacher. (Veronica, father: machine operator; mother: computer service clerk)

There was greater diversity of career choices among the Canadian women of Chinese ancestry than among the men. Those women whose choices were seen by parents as impractical or not leading to financial security (e.g., French, Asian studies, and film studies), reported conflict with parents or feelings of inadequacy, dissatisfaction, or uneasiness with self. In a few instances both university attendance and career choice for girls were seen by one or both parents as subordinate to the goal of marriage and family.

> I had originally said when I was in high school that I wanted to go into medical school and my mom said, "Oh no, girls don't do that. You'll never get a husband that way." (Lucy)

Language and Communication

Several of the young women in this sample mentioned lack of proficiency in English as preventing them from pursuing teaching as a career. Jill, who had attended high school in Canada for one year, felt teaching was a career for people who grew up in Canada:

> I have never thought of teaching as a career. Maybe it is because my language [English] is not good.

As two others put it:

> I have a problem talking in front of a lot of people, so I think I'd have to overcome that fear first. I get really nervous, usually with superiors or peers my age, but I guess with kids a little younger than I am I wouldn't be afraid but with high school level I would be. (Corrine, age on arrival: 3)

> I tend to mumble a lot, especially for words that I know I couldn't pronounce it right and I tend to get nervous ... I don't think I would be a good teacher in Canada because of my language ... I still have quite a heavy accent and I don't think it's good to be a teacher and have such an accent ... in a lot of cases, I couldn't pronounce words correctly, so I didn't choose teaching. (Sandra, age on arrival: 11)

Discrimination

The majority of this group of women of Chinese ancestry reported no experiences of discrimination in schools. Laura, however, was eloquent in supporting her view that "generally school is not a pleasant experience for minorities."

> I got into trouble in preschool, that is where I experienced my first racial remark, the first time I knew what racism was. I knew what the words were, I knew that they were bad, because my brothers and cousins would talk about it. And I met this girl and we were on a slide and she called me a chink and a Chinaman, and she was standing up on the slide and I pushed her off, and she landed flat on her face, and I got into trouble and I have been in and out of it ever since.
>
> As a child you are intimidated. There is this tall white lady telling you that you have to go to ESL, even though you watch Sesame Street and can say your numbers in French.

Josephine remarked:

> We were always wondering if teachers were wondering why we [Chinese Canadian students] were here [in French class]. Again you had to be top-notch. [Josephine had earlier explained "You have to be 110% to be accepted in the white world."] And I couldn't stand it. I'm here to learn, I'm not here to become you.

Canadian Men of Chinese Ancestry

None of these 10 men is currently preparing for or interested in a career in teaching. Only one respondent could remember ever having wanted to be a teacher. When questioned about why they were not interested in teaching careers, difficulties with language and communication skills were seen as a prominent barrier; school climate and discrimination were also mentioned by some of the respondents. Parental views and influence were identified as the background against which these other factors played themselves out.

Parental Influence

Most of these men acknowledged strong, consistent, and pervasive family influence to choose jobs from professions deemed financially secure and prestigious (such as medicine). If this was not possible, second choices were in the areas of business, computing, and accounting with teaching at the bottom of the list. These expectations are stronger for men than for women. The sons responded to these expectations in a variety of ways, generally finding means in the end to pursue their own choices. However, no one selected teaching.

> In the beginning, my parents' expectations were very high. They wanted me to go into medicine or into business. But it's a bit difficult. I found getting the grades to go into those fields difficult. But now my parents just sort of let me be, sort of. I can't get into medical school or the business faculty so right now my parents have taken a really relaxed attitude. They say, "Well do your best and then we'll see what happens from then." There used to be a lot more pressure from my parents to do certain things. (Peter, father: retired doctor; mother: homemaker)
>
> Basically, our family has always been in business so there is a little bit of pressure on going into business. (Hayne, father: investor; mother: housewife)

> [My parents] did not influence my career choice [Wesley—the only man in the sample going into medicine] and in my teen years I was sort of a rebel so I didn't really listen to my parents, but they did influence me to attend university. (father: baker; mother: laundry worker)

> There are pressures from my family to get a professional career— either a doctor or a lawyer. They felt more strongly about this before but now I know what I want so they have less of an influence. [Richard is now pursuing a business degree] ... My mother always thought business was a risky thing versus a doctor ... If you got shipped back to Germany or something, you would always have a career. (father: cook; mother: housekeeper)

As Richard points out, these parents' views of desirable occupations may be importantly formed by considerations of career security (and perhaps portability). The impact of parents' views of desirable occupations was reinforced by the financial resources that they provided to make university education possible. Regardless of parents' occupations or level of education, all but one of the interviewees said finance was not a barrier to higher education. Typical was:

> Money is not a factor because in my family an education is sacred and they will lend me the money or even borrow the money for my education ... they had a scholarship fund for me when I was a baby so I have to at least get my first degree. (Wesley, father: baker; mother: laundry worker)

All these men lived with their parents and this was noted by some as an important cost saving factor.

None of them saw their parents promoting teaching as a career of choice. Doug, the one male student in this sample who had indicated an early interest in a teaching career (but later made another choice), said:

> When I mentioned to my mother that I might want to become a teacher she questioned how good teaching would be as a career. (father: noted as deceased no occupation given; mother: teaching certificate from Hong Kong, currently a homemaker)

A number of respondents noted that their parents' views of teaching in a Canadian setting ran somewhat counter to the notion that in Chinese culture teaching is considered an honorable occupation.

> Well in Chinese circles it *used to be* highly respected to be a teacher. (Wesley)

> I always found it ironic that although, according to classic Confucian ideals, a teacher is accorded respect, in real life this hardly ever happens. (Wallace)

Language and Communication

Four of the nine men mentioned they had difficulties in the area of language and communication, which precluded their considering teaching as a career. Three of these four men had grown up in Canada. As Peter who was born in Canada put it:

> You have to have good speaking ability. You have to be able to communicate. You have to be able to express your wishes to students or whatever and you have to deal with the students on a daily basis. I don't know if I can do that.

Discrimination

Two of these men had a substantial amount to say about discrimination in educational institutions. Jason attended a multiethnic urban secondary school and noted:

> It was very bad between students. You learn to take care of yourself. Name calling is always big. Mostly between Chinese and Caucasians was the major conflict. I think I was pushed around quite a bit because I was Chinese. A scapegoat. Day-to-day things ... Teachers ... accepted it as something that happens. They didn't really do anything about it.

Wallace said that negative experiences ("social problems") in high schools made returning there for work an unattractive option:

> Minority students are not interested in teaching because of the social problems that exist in schools, with gangs, drugs, and the large immigrant population. Many minority students did not have good experiences in Canadian high schools and would not wish to go back to that same environment. Many Asian students do not integrate very well and tend to stay in their own cultural groups. Consequently they get bored and turned off and do not get very involved with the social community of the school. They are then unlikely to want to get reinvolved in schools after they graduate.

Some of these men reported no problems at all with discrimination in their schooling. Wallace was equivocal about his experiences:

> If I had had anything [discrimination] like that happen to me I would have just said to myself, got to try harder. It may very well be that I did come across that ethnic thing. If anything bad were happening in my life, my response would be try harder, try harder, try it again, overcome, be popular, try to be popular ... get on student council, don't wear plaid shirts and brown pants every day. That was my response to any sort of thing.

He then described an experience in kindergarten where other boys repeatedly stole his lunch for the entire school year.[11] He continued:

> That was one experience that pushed me to say I'm not going to take this anymore. Whatever the world throws at me I'll take it and I'll do better 'cause I have to 'cause otherwise where's my family going to go? ... That never got solved and bothered me. I decided I'm not going to let this happen to me again. I don't want any of this to happen to me or my family again.

The five men who reported that they personally had not had any difficulties with discrimination said that it was not an issue of any importance to them in making decisions about their careers.

Canadian Women of Punjabi-Sikh Ancestry

One of the seven Canadian women of Punjabi-Sikh ancestry in this sample was preparing for a career as a teacher. When the other six were questioned about their career decisions, parents, especially fathers, were perceived as influential. In general, these women did not appear to evaluate their language skills as currently inadequate to do any kind of job, although some readily acknowledged English language difficulties in the past. Concerns about discrimination were commonly expressed.

The young woman in this group who was at the time of the interview, preparing for a teaching career, was alone in citing high school teachers as encouraging in terms of education generally and interested in her career choice. She described experiences with two teachers where she had been encouraged not to quit school and to consider either law or teaching as possible careers. In the follow-up mailing this woman indicated that although she had been accepted to a teacher education program, she had finally enrolled in law school.

> By doing volunteer work with teens I found I didn't have as much patience or interest as I thought I should have had ... I was also discouraged quite a bit by old high school teachers who felt that teaching was no longer teaching but just babysitting. I must admit this had an impact on me. I began to wonder if I had what it took to avoid adopting such a pessimistic view of teaching. (Pindy)

Parental Influence

All these young women indicated that their parents' opinions were of concern to them. Six of the seven women cited their fathers as the primary influence on their career decisions. The seventh woman's father is deceased. One woman put it this way:

> I get a lot of my ideals from my dad, you know, this idea that I have a responsibility toward community and to put something back and to help change people's lives. I am very politically committed and that comes directly from my father. (Rajinder, father: social worker; mother: home manager)

Another woman studying medicine said:

> My dad's encouragement was always positive ... I always wanted to make my dad proud of me because I am really close to him. (Surjeet, father: mill worker; mother: homemaker)

These women singled out their fathers as influential in terms of their career choices, often remarking that their mothers were more concerned with their "happiness." As one put it:

> My father has always been the one to encourage me to do well. Always since I was 6 or 7 years old. My mother was always happy if I did well and it was important, and if I didn't I would hear from her, but I talked to my dad when it came to school. The general idea of having done well and getting a good job in the end was important. There was definitely a difference. My mom would want me to be happy and my dad would want me to be successful. I was talking to my cousin and she said almost the same thing. I don't know if it is a general female versus male stuff. (Jathinder, father: MA degree in math, works as a bus driver; mother: has a teaching certificate, works as a secretary)

Three of the seven women indicated that their parents would find teaching an acceptable career for them. In one case the daughter indicated she thought they would be quite pleased with this choice and related that both parents had been teachers in India but were now working respectively as a secretary and a bus driver. One indicated that her parents would like it better than police work (which she is pursuing) because "they would think it would be safer." The third said that teaching is highly respected, but that "the money wasn't that good."

Language and Communication

Three of the women in this sample mentioned English language difficulties when they first came to Canada, or when they started school. Kuldip (now 22), who came to Canada at age 14, reported that she had initially had difficulty acquiring her first (part-time) jobs due to language barriers. After some time, however, she felt no further difficulties in this area and did not mention any difficulties with English as being involved at all in her career choices.

Surjeet, born in Canada with Punjabi as a first language, describes difficulties when she first started school:

> I used to cry and cry. I was very shy. I couldn't speak English and I went to this humongous school. I was always nervous about my English. There were new immigrants and I wouldn't talk to them because I felt they just got here and it wasn't cool.

Surjeet is now enrolled in medicine, and English language proficiency was not mentioned as a factor in her career choice.

Sarita (now 20) immigrated to Canada as an infant and described her first languages as Punjabi and English. Her mention of language is ambiguous:

> After grade 12 I was very scared as to where I was going to go. And I didn't have the language so I had to go to college.

After deciding that the college courses were not of interest, Sarita enrolled in criminology at university and mentions no further difficulties with language.

Discrimination

In the questionnaires only one woman wrote about experiences with racial discrimination, but in the interviews six of the seven women spoke about these matters. These young women reported that the most overt hostility they encountered was from other students rather than teachers. However, teachers rarely intervened when they saw racist events between students. The following excerpts from interview transcripts show that they perceived discrimination occurring across the grade levels in a variety of ways and that coping strategies were various as well.

> Students were very cruel ... The teachers were OK ... They were supportive. They went out of their way ... Things have changed. When they came here they [Sikh parents] didn't have the facilities. They all came from villages. They had to assimilate, they had no choice. They had to bring up their children this way. My mother felt like she had to wear a miniskirt, otherwise people would harass her walking down the street in a sari. It is a double whammy if you are adolescent and a minority. (Surjeet)

> I remember one [high school] teacher saying, "You can't see her blush because she is so dark." And you don't know how to take it because this person is a teacher and they're laughing and it's a joke ... [But] most teachers were fairly good. (Rajinder)

Rajinder also expressed interest in education, but not in a public school setting because she felt intolerant of the racism of youngsters:

> The kids in the education system at the moment are very obnoxious. Maybe not all of them and in different schools it might vary, but I just had this experience working with a group of kids for a few weeks and they were so racist, I couldn't believe it ... Actually there was one person in the group who was also Indian and they would make comments to him, and then they would make sort of oblique comments to me about the color brown. It was just obvious from the tone of their voice that they were being racist and I couldn't deal with it at that point and I really didn't want to.

Jathinder spoke about discrimination in this way:

> There were incidents in grade 3 but it never affected me in any way although it was to people of Indian origin. It didn't bother me. It never happened in front of the teacher so you don't go running for help. I just thought they were really silly and immature. Nothing serious ... One thing I remember from high school was from a teacher we all hated, everybody. Generally people consider me to be a shy person or quiet, so I don't know whether I answer something loud enough. This teacher said to me, "You don't have to be so quiet because you are Indian. Maybe it is expected of you in your culture but ..." I don't think of myself that way. A lot of people think Indian women are submissive and I am not like that and neither is my mother. I don't know if he was trying to be rude, we all thought he was a jerk anyway. In university in commerce ... you have to take on Caucasian values to succeed ... It is nothing the teacher says or anything but institutionalized racism. It may also have to do with gender.

Harjeet shows her strategy for avoiding racism:

> I experienced less racism in grades 11 and 12 because I didn't hang around with other East Indians. When I was in grade 7, we had a girl who came from India and they asked me to help her out, my being East Indian and everything. That's what they have ESL teachers for ... I don't think students should have to bear the burden of teaching another student a new language ... Kids would say or imply that I should stay with "your East Indian friend."

Rajinder spoke of her impatience with a university class in which she felt the professor and other students were avoiding discussion of clearly evident racism:

> They don't talk about realities in a lot of arts classes. I just had an English class last semester and we were discussing some very political texts, and I watched and no one said the word racism for three weeks. That really frustrated me because obviously that's what these texts were about. Toni Morrison, Alice Walker, that's what these people talk about. (Rajinder)

Canadian Men of Punjabi-Sikh Ancestry

One of the five men interviewed here is preparing for a career in education. In choosing a career Rick's first motivation was to earn a lot of money, so he entered the Faculty of Business, but he soon realized based on his experiences teaching soccer and assisting in a day care center that he valued other things, such as working with children. His parents were happy with his choice.

The men who were not interested in teaching as careers had made rather detailed plans for their futures, involving short- and long-term goals. Although questioned specifically, only one of these men gave a clear reason for not choosing education, which was his father's discouragement. These men did not indicate that they felt any personal inadequacies had precluded their choosing

teaching as a career: only one, for example, mentioned English language proficiency as a personal problem, but one he had solved. The sense one has from reading transcripts is that these young men have simply made other choices.

Parental Influences

Four of these five men, like the women, express the importance of their fathers' wishes in their career choices.[12]

> I want to be something that would be respectable in my dad's eyes. My dad wants me to become an accountant because he thinks that is a good job to have. He just wants me to graduate with a good degree so he can say "my son is graduating with so and so." There is a bit of pressure but not a lot. My mother doesn't think about it too much. She just wants me to graduate. My dad really talks to a lot of people and he knows my personality really well ... So I go to my dad for a lot of advice. (Rick, father: assistant engineer; mother: chef's helper)

Language and Communication

None of these men cited language difficulties as being implicated in their career decisions.

Discrimination

With regard to issues of prejudice and discrimination, these men had a variety of perceptions.

> My whole life I have hung around with English people. I find if I have a white friend I can trust him more, because it tends to be that a lot of East Indian people if they don't like you they will start rumors about you. There is a lot of jealousy. I had no problems in my school because I knew a lot of English people, so I was always treated like one of the guys. In elementary there was not a great understanding of religion but it didn't matter. Some kids you get bugged by just because you are different. I was glad I didn't have to wear a turban ... I was very lucky. I have great parents. They're very westernized. My parents can talk to my friends. (Rick)

Anoop said about discrimination:

> I just think there is no communication there [if minority students are treated unfairly in school]. If I have a problem, I go up and tell the teacher. I say, "Listen, why is this the case? Why am I doing poorly?" and usually the teachers are very helpful in that sense.

He did comment on racial differences and teaching, however:

> I've always associated teaching as a white profession.

Perry (whose father is a high school teacher) dealt with the question of personal experiences with discrimination by talking about what he thought teachers might do to alleviate such difficulties:

> I think that it should be a teacher's responsibility, especially if they are teaching in a multicultural district, to understand the cultures a little bit more ... it goes beyond the significance of why they wear a certain type of clothing or their food and dancing; religion is a really touchy subject, but it is such a fundamental subject that I believe that kids should be exposed to why people's beliefs are the way they are, it just leads to a bit more understanding, teachers have to play a bigger role. I wouldn't limit that to ethnic groups, I would say it's just social issues. Making sure that teachers treat minority students well while they are in

grade school is an important first step in encouraging minority students to become teachers.

Discussion

Parental Influence

These students' reports as a whole indicate that their parents' play two important roles in their career choices: one is financial; the other relates to views of the desirability of particular careers. Although some of the students had to work or live at home in order to attend university, almost all indicated that their parents' support for postsecondary education was such a high priority that they would find the money for their children's education regardless of their socioeconomic circumstances. For these groups the relationship of social class and educational attainment is different from that found in earlier large-scale Canadian surveys showing that students from families of higher socio-economic status are much more likely to attend university than their peers from lower socioeconomic backgrounds (Anisef & Okihiro, 1982; Porter, 1965; Porter, Porter, & Blishen, 1979).

Parents also play an important role in selection of specific careers. For Canadians of Punjabi descent it appears that parents' views, especially their fathers', are important in decision making. The greater influence of fathers in determining educational goals and choosing a career is also identified by Wakil, Siddique, and Wakil (1981) in their study of Indian and Pakistani families in a western Canadian city.[13] The students we interviewed of Punjabi-Sikh ancestry indicated that their parents would have no strong objections to teaching as a career choice. In the cases of the one female and one male who had chosen it, there was parental support. The males of Chinese ancestry reported their parents would not be supportive of a decision to enter teacher education. In the case of the females of Chinese ancestry, eight of the 12 women reported parental approval of teaching as a career choice for them, but only one of these young women has chosen this path. The one woman of Chinese ancestry who was preparing for a teaching career reported her parents were *not* supportive of this aspiration (nor had her teachers been encouraging in this regard). In general, then, the parents of students of Chinese ancestry are not supportive of teaching as a career for their sons, but are less unified in their opposition to this career for their daughters. Nevertheless, these young women are generally not choosing teaching as a career for themselves, sometimes because, they report, of inadequate English language skills. This matter of language skills is discussed below.

Language and Communication

The careers identified in our study as desirable to males and females of Chinese ancestry are also those where quantitative technical skills take priority over linguistic or communication skills. Some of these students are native-born, whereas others were older on arrival in Canada. The predominant number of students of both Chinese and Punjabi-Sikh ancestry grew up in families where the English language played a minor role. Despite many years of English medium schooling, several cite English language difficulties as reasons for not choosing teaching (or other careers for which verbal skills in English are necessary). These difficulties with language are mentioned often in American

literature concerning students of Chinese ancestry. Pang (1990) notes with regard to these individuals, "They not only feel the inability to do well but reveal a fear of writing and speaking" (p. 58). Hsia (1988) states that the contrast between Asian Americans' achievement in quantitative fields and their avoidance of the difficulties in the fields that demand well-developed verbal skills is stark among recent immigrants and is still noticeable, even after several generations among the native born (Chinn & Wong, 1992). Chinn and Wong argue that overlooking the relatively weak verbal skills of Asian students and making no educational provisions for improving those skills is unfortunate because

> Early development of these skills can enhance social interaction with non-Asian peers and reduce the avoidance of courses out of the quantitative areas. The development of verbal skills, coupled with the acquisition of social skills, can enhance the likelihood of Asians branching out from their traditional areas of study and thus help in their recruitment into teacher education programs. (p. 131)

Unlike the Canadian students of Chinese ancestry, none of the students of Punjabi ancestry mentioned current difficulties with English as responsible for their decisions not to pursue teaching careers. Some acknowledged previous difficulties, but all appeared to regard these as difficulties they had overcome that were no longer relevant.

Discrimination

Some of the Canadian students of Chinese ancestry spoke about experiencing discrimination in their own educational experiences, and two implied that racial, as well as other tensions in the Canadian public school system made them unattractive work places for minority students. Although some clearly and strongly articulated difficulties they had had as students in schools, others had little to say in this regard and said it was not an issue of any importance to them in career decision making. The American literature on students of Asian ancestry stresses the need for schools to be sensitive to their emotional and social needs (Pang, 1990). The point is made that students who feel powerless to deal with prejudicial situations may withdraw from the school community, thus diminishing the likelihood of pursuing careers in teaching.

The Canadian students of Punjabi descent were also diverse in their responses to questions about prejudice and discrimination. Five of the six women interviewed described incidents of racism in their own schooling, and one indicated that the racism of students in the school system was part of why she was not personally interested in a teaching career. The Canadian Punjabi men had less to say about these matters. They reported they had personally dealt with racial discrimination by ignoring it or by avoiding identification with their own racial group. These men displayed little interest in discussing these matters.[14]

Interpretation and Recommendations

We began this investigation reminded by McCarthy (1988, 1990) that diversity in the experiences within and between minority groups needs to be recognized and investigated. In looking in detail at the experiences of individuals from the two most populous groups in British Columbia (and looking at them by

gender), we were expecting diversity. This expectation was not unfounded. We were continually struck in reading the interviews and listening to the tapes by the many ways individuals made decisions and managed problems as they set about their task of getting the education they needed to make a living. At the same time, we also noticed commonalities in the experiences and perceptions of members within each of the four groups described. How does the information we have gathered and our analysis of it affect our perceptions of equity policies and barriers to the representation of Canadians of Chinese and Punjabi ancestry in the field of teacher education? How can we improve? Our analysis has identified three main dimensions: parental influence, language and communication, and discrimination. We discuss each in turn below.

Parental Influence

For both groups studied here, reaching out to parents is important, but it appears that what parents need to know will vary somewhat between the two groups. When the men of Chinese ancestry were asked what could be done to encourage more minority students to go into teaching, suggestions focused on prestige and finance:

> You have to influence the parents to change their ideas and also to have better financial incentives. (Peter)
>
> Sell the notion of giving it a prestigious status. (Wallace)
>
> Enhance the perception of teaching in the eyes of Chinese parents. (Doug)

Although faculties of education clearly will not in the foreseeable future affect the remuneration offered to teachers, it is apparent that if recruitment into faculties of education from these minority communities is desired, communication with parents will be important. Outreach to minority community organizations and through the ethnic media are avenues that have not yet been tried but that might yield positive results. Media and public information targeted at communities should be available not only in English, but in Cantonese, Mandarin, and Punjabi and should include perspectives on these careers not only from students, but from parents and respected community leaders including those who have successfully pursued careers in teaching. We are aware that many outreach efforts to the community at large focus primarily on the personal satisfaction and fulfillment associated with a career in teaching. In view of differential concerns between and among groups identified in this research, it would be important to be specific about a wider range of occupational concerns including socioeconomic status (e.g., financial security and prestige) as well as personal esteem, satisfaction, fulfillment, safety, and service to the community.

As our research has indicated, there is fairly strong resistance to teaching as an appropriate career choice for men of Chinese ancestry, and so attention will need to be paid to this gender issue in recruitment activities. In addition, because there seems less resistance to women going into teaching careers, it may be that recruitment of women of Chinese ancestry will be less problematic than in the case of men.

In addition to the above considerations statements by mainstream and minority educators should deal forthrightly with issues of language acquisition and discrimination in the schools.

English language proficiency may be regarded (as many of the Canadian students of Chinese ancestry see it) as a personal problem or a "cultural" pattern. Students of both Punjabi and Chinese ancestry identified English language acquisition as problematic, with only the Chinese seeing this as persistent. As indicated by the data presented earlier, difficulties were experienced by those who emigrated in their teenage years, but also by several who were born in Canada.[15] Although this research does not inquire into the reasons for the differences in persistence of problems between the two groups, it would seem that linguistic as well as social factors may account for the variance.[16]

It is important also to remember that career decisions have not only to do with differential socialization, talents or interests, but also with opportunities and institutional arrangements. Seeing Canadian males and females of Chinese ancestry as being "naturally" and "culturally" concerned with science and business, and absent from arenas where work is importantly conducted through oral communication, for example, may ignore structural arrangements that make it difficult for these young people to enter these professions. What they identify as linguistic barriers might encompass what they also perceive as social barriers. This research has clearly shown that a number of individuals from at least this group of Canadian students of Chinese ancestry self-evaluate themselves as relatively unskilled verbally and unable to take on verbally oriented careers. Are the students right: are their verbal skills really less developed than would be necessary for jobs like teaching? Why would a group of students who have lived in Canada for many years, attended English-medium schools, and have enough intellectual ability to attend university not have developed correspondingly in English language proficiency? Have individuals from this particular background been exluded from certain kinds of conversations? Have they been somehow persuaded that their English language abilities are inadequate? Questions like this suggest that what might appear to be personal and individual inadequacies may be constructed impediments to the participation of individuals from particular groups in particular jobs. This structural analysis is supported by Li (1988a) who identifies relative absence of Canadians of Chinese ancestry in jobs requiring social interaction as a strategy for avoiding disadvantageous competition.

Distinguishing between cultural and structural sources of lack of participation may be an intellectual nicety of little interest to the students concerned. If high schools wish to increase the options in career choices of all students, it may be that English language courses need to be available more widely than heretofore. More emphasis on oral communication through drama and role-play would be beneficial in preparing for a teaching career. If faculties of education are to encourage more minority students to enter and complete teacher education, it may be that those faculties need to offer ways for some of those students to improve their English language skills or to increase their confidence in speaking in certain arenas. As we prepare anglophone teachers to work in French immersion settings, we offer courses explicitly designed to improve candidates' French language proficiency. If universities are to prepare

minority students to become teachers, students may need opportunities to improve their English language skills so that this job requirement does not preclude their participation.

Discrimination

Clearly individuals in each of the four groups have experienced discrimination differently, and have learned to regard and deal with it differently. The Punjabi women students appeared to be most acutely aware of it. (Our research was not focused on reasons for variation in articulations of discrimination). They may relate to real differences in discrimination experienced by the two groups or to different levels of awareness. It is difficult objectively to relate and compare the severity of discrimination, but at least in recent years the Canadians of Punjabi-Sikh ancestry have been the focus of much media attention in this area.[17] Many of the students have articulated their perceptions of secondary school incidents of discrimination on the basis of race or ethnic identification. In some cases, students said that they perceived schools to be unwelcoming to "people like them," and that it was because of this they had no wish to work in such places. None of the students we interviewed had participated in multicultural leadership camps that are available (on a limited basis) in their schools, and none spoke of multicultural programs or curriculum. The one student who mentioned the positive influence of a minority (teacher) role model was referring to his father.[18]

Not all the students we interviewed experienced overt instances of personal discrimination. In our view all experienced implicit institutional discrimination in schools that took little notice of their unique positions. The school lives of these students appeared to be untouched by the variety of multicultural and antiracist programs and curricula that have a "social change mission," and "link power and empowerment with race, social class, and gender issues" (Sleeter, 1991, p. 2). They could not refer to any examples in their experience of the four key areas for social transformation of schooling and empowerment of students identified by Cummins (1986).[19] In spite of this, they were successful in the conventional norms of educational institutions. But we see their reluctance to pursue teaching as a measure of the ways these institutions implicitly exclude/deny and discriminate against them. Clearly, if minority students are not to be discouraged from working in schools as teachers, the climate in public schools for minority students needs to change.

In spite of the apparent absence of contact with multicultural/antiracist programs, the students we interviewed had many ideas about what teachers could do to improve the climate in high schools. For example,

> Make others aware that there are other cultures here ... I guess they [other people] find that some of the things we do are weird or different and they don't understand why. (Veronica)

> If students feel differences ... they need to feel safe enough to talk ... but not to have the issue blown up into something major, but to have it dealt with ... When teachers try to be multicultural, it can make people proud or it can make them more self-conscious. You don't want to be made to feel self-conscious and different. But on the other hand you don't want to be ignored. It is very sensitive, it's very hard not to do it without generalizing and flattening, it takes a lot of commitment from the teacher. (Rajinder)

Because much of the responsibility for school change falls to teachers, the absence of comprehensive programming in multicultural/antiracist teacher education is troubling and problematic (Beynon & Warsh, 1993; McCarthy, 1990; Sleeter, 1992). This area is discussed in the concluding section.

Conclusion

We agree with Fleras and Elliott (1992) and Henry and Jain (1991) that employment equity legislation directed at the hiring process is important. However, this legislation (and what appears to be the inevitable resistance to it) must not divert attention from the necessity of initiating programs for educational/training equity that are attentive to diversity among and within groups and consequently provide students with appropriate preparatory skills and experience.[20]

In this research we have been guided by the interactionist emphasis on ascertaining the meanings that guide individual behavior and McCarthy's (1988, 1990) emphasis on diversity between and within ethnic groups. These have been useful in providing insights into the ways social institutions and cultural influences impinge on individual perceptions and actions regarding the choice of teaching as a career. We have seen that at least two prominent visible minority groups have some commonalities in regard to parental influence, but also that they are different in important ways (e.g., their experiences of discrimination and their self-assessment of their skills in English language and communication.) As policy analysts and implementers we see the knowledge of the group and the search for commonalities serving as a heuristic in identifying where and why inequities may exist for visible minorities vis-à-vis mainstream Canadian institutions. As educators we are acutely aware of the need to view a group in interactionist terms through the descriptions, meanings, and explanations provided by individual students. Knowledge of the individuals will help both in preventing the group perspective from becoming a stereotype and in designing educational initiatives that will respect the needs and concerns of the individual students.

Thus as policy analysts, implementers, and educators we need to become adept at using analytical bifocals so that we can shift easily from a group to an individual perspective. With this general approach in mind we turned to specific recommendations for improving the representation of minorities in teacher education. Even with this small-scale study we see individuals who, in spite of institutional barriers such as inadequate language training and garden variety racism, *do* wish to pursue training or employment in teaching. The support, encouragement, and outreach of mainstream educational organizations such as the teachers' federations, universities, and ministries of education should play a stronger role in assisting them, and other minorities as well.[21]

Our research also addresses some practical issues related to making closer links between the world of school and the world of work. Several provincial governments and ministries of education in Canada are embarking on renewed efforts to have students develop work-related skills prior to entering postsecondary education. In British Columbia the Ministry of Skills Education and Training (formerly Advanced Education) and the Ministry of Education have joined forces in launching the Skills Now initiative, which builds career and personal planning and work experience into the curriculum in increasing

percentages from grade 4 through to graduation. These initiatives acknowledge the necessity of a partnership with parents, but are otherwise vague as to what this should entail. Our research indicates that, at least as far as teacher education is concerned, these partnerships might need to be as varied as the parent groups to whom they are reaching.

Our research also supports the intent of both ministries to provide youth with positive school experiences, including mentorships and internships that support an informed process of career selection. New efforts in this area need to take special care that students from all groups are reached and that this may require unique approaches including collaboration with community-based organizations and youth groups that generally operate outside school boundaries.[22] Such efforts are necessary in order to provide young people convincing evidence that they are valued both as students and as future members of the teaching profession.

Although we view these government initiatives as potentially useful and supportive of minority students, we also fear that they may have little effect unless pervasive concerns about school climate are addressed. As long as this climate does not actively support students through multicultural and antiracist as well as other kinds of programming; as long as teachers are not routinely prepared by universities and professional associations to take an active role in transforming their classrooms, we doubt that new government programs will reach all students and provide them with equal opportunities.

Notes

1. Concern about equity of access and equality of representation in various fields of employment has been evident in Canadian society at least since the Royal Commission Report on equality in employment (Abella, 1984) and the subsequent passage of the federal Employment Equity Act (Government of Canada, 1986). Universities and other federally funded employers, for example, are required to review the composition of their work force with a view to determine whether it is representative of First Nations, visible minorities, women, and persons with disabilities; if it is not representative these institutions are required to develop plans to address this imbalance in future hiring.
2. These estimates of representation are based on averages of 10 semesters between 1986 and 1990. During this period the representation of Canadians of Chinese ancestry was consistently low, whereas in the case of the Canadians of Punjabi ancestry there was improvement approaching equity in the more recent four semesters. Determining whether this improvement will be sustained requires ongoing monitoring. For more specific data on this topic see Beynon et al. (1993).
3. Our first study focused on an analysis of data from questionnaires that asked respondents to rate the relative influence of 12 factors influencing their choice of career. These factors relate to key intrinsic, extrinsic, and interpersonal influences identified by research among the general population into the choice of teaching as a career (Carpenter & Foster, 1977). These factors were: interest in subject matter; opportunity to do work that is personally satisfying; opportunity to do work that helps others (intrinsic); availability of jobs with good salaries; previous employment or volunteer experience (extrinsic); inspiration or encouragement by someone already practicing the career; family members' expectations; parents' wishes for the future; inspiration or encouragement by a teacher or counselor; ethnic group expectations; career choices of friends; and inspiration or encouragement by a guru, minister, rabbi, or priest (interpersonal). Statistically significant differences between ethnic groups occurred only in regard to three interpersonal factors: family members' expectations, parents' wishes and ethnic group expectations. The overall means for these factors were relatively low for all groups (compared with intrinsic and extrinsic factors); but the influence on the minority group individuals was significantly greater than on the Canadians of Anglo-European ancestry.

4. Using 1981 census data Li (1988a) demonstrates that 75% of Canadians of Chinese ancestry were foreign born and over 75% indicated the Chinese language as their mother tongue. This apparent broad similarity does not, however, reveal specific languages or countries of origin.
5. Investigating in detail the similarities as well as contradictions and tensions within and between groups requires historical research, and we know that the two groups focused on here have a daunting variety of historical experiences in immigrating to and residing in this province. This study does not examine many facets of their variation, although we are certain that those differences will be important in characterizing the experiences of individuals and groups.
6. The questionnaire was broadly focused on the myriad factors that could influence career choice: parents, peers, mentors, salaries, critical events, interests, barriers, and so forth. The interviews were guided by open-ended questions about selection of academic program and career. Students were asked to describe their current programs, how they came to be in these, what careers they were planning to go into, the relationship of their family's wishes to their plans. They were asked about their own and their family's and ethnic group's view of being a teacher, and whether there were particular individuals who were influential in guiding and supporting their academic and career choices. In an attempt to ascertain whether their own schooling experiences affected their perceptions of teaching as a career, they were asked whether there were times when they felt all students weren't treated equally, what the reasons for this were, and what if anything teachers could do to meet the needs of students from all groups. Finally, students were asked whether they thought all ethnic groups were equally represented in all careers, what might account for levels of representation (as they saw it) and whether they had any suggestions about ways of increasing the involvement of minorities in teaching.
7. The classification of Canadian occupations used by Li (1988b) and adapted from Wright (1979) is: employers, managers, professionals, petty bourgeoisie and workers.
8. Our findings in the previous study were in concert with Li's (1988 a) analysis of 1981 census data for Canadians of Chinese ancestry showing that about 29% of Canadians of Chinese ancestry had at least some university, compared with about 16% of the rest of the population.
9. Perhaps our study of secondary school students that includes much larger numbers of students who had not yet faced decisions about university attendance would reveal a different picture about the relationship between parent occupation, access to postsecondary education, and desirability of teaching as a career.
10. For students of Chinese ancestry parents' places of birth were almost equally divided between rural China and urban Hong Kong. Eighteen of 22 students identified these as parents' place of birth. However, most of these parents had spent significant parts of their lives in Hong Kong before emigrating to Canada. Of the four remaining students two identified Indonesia, one Vietnam, and one Singapore as parents' (or own) country of origin. All students of Punjabi-Sikh ancestry identified the rural Indian province of Punjab as their parents' place of birth.
11. In a written response to the transcript, Wallace wondered if the major motivation for this incident had been racism or if it had been a matter of older, bigger children bullying a smaller child.
12. Jagdish, who was 33 years old at the time of the interview, was an anomaly in not finding his parents' wishes a major factor in his decision making; we expect that this may have been related to his age.
13. In this study of 50 families only eight were of Sikh background.
14. Gibson and Bhachu (1991) describe a "climate of prejudice permeat[ing] the school experience of all Sikh students" in the farming community in northern California that they studied, such that the Sikh students are "verbally and physically abused by majority students, who refuse to sit with them in class or on buses, crowd in front of them in lines, spit at them, stick them with pins, throw food at them and worse" (p. 69). Like the Canadian students with whom we have been working, these American Sikhs, despite their difficulties, were completing high school, and the boys (at least) were aspiring to postsecondary education.
15. We did not on the basis of our interviews make judgments about language proficiency. As researchers and teachers we are quite accustomed to variations in spoken English and are unlikely to make the same judgments about accent as others might. We are also aware of the research that indicates that oral fluency can be achieved in a significantly shorter period of time (2 to 3 years) than can proficiency in writing (5-7 years) (Cummins, 1986). Thus for the

purposes of this analysis we only considered the students' self-perceptions of their language proficiency.

16. Punjabi as an Indo-European language is more similar to English than is Chinese and language similarities may well be important in ease and speed of acquisition.
17. The national controversy about turbans and the RCMP, as well as admission to Canadian Legion halls has put Sikhs in the spotlight in a way that is not apparent in the case of the Chinese. In general, in the Punjabi-Sikh community modes of dress are an important part of cultural and religious practice. For secondary school students where peer dress norms are of tremendous importance, it could be harder for Punjabi than for Chinese students to fit in. The Punjabi experience of ethnic conflict in their country of origin could also be a salient factor in creating sensitivity to and awareness of discrimination in Canada. Although there is certainly a long history of discrimination against both Chinese and Punjabis in Canada, it is not immediately apparent whether most of the Chinese who responded in this study were unaware of or simply unwilling to discuss this matter with our interviewers.
18. A number of districts in the greater Vancouver area have multicultural/antiracist programs but they are available on a limited basis to selected and self-selected individuals. (Fisher & Echols, 1989). Similarly there are some visible minority teachers in these schools but the students we interviewed had not had any classes with them.
19. These include incorporation of students' language and culture, collaborative community participation, instructional strategies that draw students into participation, and assessment procedures sensitive to their linguistic and cultural heritage.
20. Failing provision of these programs, employers can continue to claim that they would hire minority applicants as required by legislation but that there is a dearth of qualified applicants. In the case of educational authorities (school boards and teacher education institutions) it is then possible to resort to the explanation that a particular minority group is simply unsuited to the profession or that parents and students alike are focused on more prestigious and financially rewarding occupations. These are explanations we commonly hear in institutional efforts concerning underrepresentation. We are also aware of the concern many individual minority students have expressed that employment equity legislation singles them out for special treatment simply because of their minority status and not because of their skills and abilities.
21. Each of these organizations to a greater or lesser extent is doing some work in this area. We do not mean to belittle the importance of the work done by educators committed to transformative education. The problem is that their work is marginalized and encapsulated in a committee, a course, or a department and does not pervade the fundamental operating assumptions of the organization.
22. A number of the women of Punjabi ancestry noted that there was a Punjabi women's association they belonged to that was very active in mentoring programs, peer counseling, and bridging communication between parents and young women. They indicated that because many young women did not attend temple, this group could be more effective in reaching a wider group of young women.

Acknowledgments

Funding for this study was provided by the SSHRC and by the Faculty of Education and President's Research Grant of Simon Fraser University, Burnaby, British Columbia. We acknowledge Donna Clark, Shemina Hirji, Kaui Keliipio, and Sherry Roberts for their work in interviewing and all the students who generously volunteered to be interviewed. For insightful editorial assistance we thank Dr. Margaret Gold.

References

Abella, R. (1984). *Equality in employment: A royal commission report*. Ottawa: Minister of Supply and Services Canada.

Anisef, P., & Okihiro, N. (1982). *Losers and winners*. Toronto, ON: Butterworth.

Banks, J. (1991). A curriculum for empowerment, action and change. In C. Sleeter (Ed.), *Empowerment through multicultural education* (pp. 125-141). New York: SUNY Press.

Bannerji, H. (Ed.). (1993). *Returning the gaze: Essays on racism, feminism and politics*. Toronto, ON: Sister Vision Press.

Baptiste, M.L., Baptiste, H.P., & Gollnick (Eds.). (1980). *Multicultural teacher education: Preparing educators to provide educational equity*. Washington, DC: American Association of Colleges for Teacher Education.

Bennett, C. (1988). Assessing teachers' abilities for educating multicultural students: The need for conceptual models in teacher education. In C.A. Heid (Ed.), *Multicultural education: Knowledge and perceptions* (pp. 23-38). (ERIC Document Reproduction Service No.ED 312-196)

Beynon, J., Toohey, K., & Kishor, N. (1993). Do visible minorities in British Columbia want careers in teaching? *Canadian Ethnic Studies, 24*(3), 145-166.

Beynon, J., & Warsh, M. (1993). Multicultural/anti-racist education: Policy to practice. *Research Forum, 11*, 28-33.

Breton, R. (1970). Academic stratification in secondary schools and educational plans of students. *Canadian Review of Sociology and Anthropology, 7*, 17-34.

Buttrick, J. (1977). *Who goes to university from Toronto?* Toronto, ON: Ontario Economic Council.

Carpenter, P., & Foster, B. (1977). The career decisions of student teachers. *Educational Research and Perspectives, 4*(1), 23-33.

Carty, L., & Brand, D. (1993). Visible minority women: A creation of the Canadian state. In H. Bannerji (Ed.), *Returning the gaze: Essays on racism, feminism and politics* (pp. 169-182). Toronto, ON: Sister Vision Press.

Chinn, P., & Wong, G.Y. (1992). Recruiting and retaining Asian/Pacific American teachers. In M.E. Dilworth (Ed.), *Diversity in teacher education: New expectations* (pp. 112-133). San Francisco, CA: Jossey-Bass.

Clifford, J. (1988). *The predicament of culture: Twentieth century ethnography, literature and art.* Cambridge, MA: Harvard University Press.

Cummins, J. (1986). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56*, 18-36.

Fisher, D., & Echols, F. (1989). *Evaluation report on the Vancouver School Board's race relations policy.* Vancouver, BC: Vancouver School Board.

Fleras, A., & Elliot, D.L. (1992). *Multiculturalism in Canada: The challenge of diversity.* Scarborough, ON: Nelson Canada.

Gibson, M.A., & Bhachu, P.K. (1991). The dynamics of educational decision making: A comparative study of Sikhs in Britain and the United States. In M.A. Gibson & J.U. Ogbu (Eds.), *Minority status and schooling: A comparative study of immigrant and involuntary minorities* (pp. 63-95). New York: Garland.

Gilbert, S. & McRoberts, H. (1977). Academic stratification and educational plans: A reassessment. *Canadian Review of Sociology and Anthropology, 14*, 34-46.

Gollnick, D.M., Osayande, K., & Levy, J. (1980). *Multicultural teacher education: Case studies of thirteen programs.* Washington, DC: American Association of Colleges for Teacher Education.

Government of Canada. (1986). *Employment equity act.* Ottawa: Author.

Grant, C.A. (Ed.). (1983). Multicultural teacher education—Renewing the discussion: A response to Martin Haberman. *Journal of Teacher Education, 34*, 29-32.

Hale, S. (1991). Feminist method, process and self-criticism: Interviewing Sudanese women. In S. Berger Gluck & D. Patai (Eds.), *Women's words: The feminist practice of oral history* (pp. 121-136). New York: Routledge.

Hammersley, M., & Atkinson, P. (1991). *Ethnography: Principles in practice.* London: Routledge.

Henry, F., & Jain (1991, April 12). When inequality is built right into the system. *The Globe and Mail.*

Hsia, J. (1988). *Asian Americans in higher education and at work.* Hillsdale, NJ: Erlbaum.

Le Compte, M.D., & Preissle, J. (1993). *Ethnography and qualitative design in educational research.* San Diego, CA: Academic Press.

Li, P.S. (1988a). *The Chinese in Canada.* Toronto, ON: Oxford University Press.

Li, P.S. (1988b). *Ethnic inequality in a class society.* Toronto, ON: Wall and Thompson.

Lynch, J. (1986). An initial typology of perspectives on staff development for multicultural teacher education. In S. Modgil, G. Verma, K. Mallick, & C. Modgil (Eds.), *Multicultural education: The interminable debate* (pp. 149-165). London: Falmer.

Martin, R. (1991). The power to empower: Multicultural education for student-teachers. In C. Sleeter (Ed.), *Empowerment through multicultural education* (pp. 287-298). New York: SUNY Press.

McCarthy, C. (1988). Rethinking liberal and radical perspectives on racial inequality in schooling: Making the case for nonsynchrony. *Harvard Educational Review 58*, 265-279.

McCarthy, C. (1990). *Race and curriculum.* Bristol, PA: Falmer.

Ng, R. (1993). Racism, sexism and nation building in Canada. In C. McCarthy & W. Crichlow (Eds.), *Race, identity and representation in education* (pp. 50-59). New York: Routledge.

Ogbu, J.U. (1991). Immigrant and involuntary minorities in comparative perspective. In M.A. Gibson & J.U. Ogbu (Eds.), *Minority status and schooling* (pp. 3-33). New York: Garland.

Pang, V.O. (1990). Asian-American children: A diverse population. *Educational Forum, 55*(1), 49-66.

Pang, V.O. (1991). Teaching children about social issues: Kidpower. In C. Sleeter (Ed.), *Empowerment through multicultural education* (pp. 179-198). New York: SUNY Press.

Peng, S.S. (1985). *Enrollment patterns of Asian-American students in postsecondary education.* Paper presented at the annual meeting of the American Education Research Association, Chicago.

Porter, J. (1965). *The vertical mosaic.* Toronto, ON: University of Toronto Press.

Porter, M., Porter, J., & Blishen, B. (1979). *Does money matter? Prospects of higher education in Ontario.* Toronto, ON: Macmillan.

Ruiz, R. (1991).The empowerment of language-minority students. In C. Sleeter (Ed.), *Empowerment through multicultural education* (pp. 217-228). New York: SUNY Press.

Singh Ghuman, P.A. (1993). *Coping with two cultures: British Asian and Indo-Canadian adolescents.* Clevedon, UK: Multilingual Matters.

Sleeter, C. (1991). Multicultural education and empowerment. In C. Sleeter (Ed.), *Empowerment through multicultural education* (pp. 1-23). New York: SUNY Press.

Sleeter, C. (1992). *Keepers of the American dream: A study of staff development and multicultural education.* Washington, DC: Falmer.

Stacey, J. (1991). Can there be a feminist ethnography? In S. Berger Gluck & D. Patai (Eds.), *Women's words: The feminist practice of oral history* (pp. 111-120). New York: Routledge.

Van Maanen, J. (1988). *Tales of the field: On writing ethnography.* Chicago, IL: University of Chicago Press.

Vetter, B.M., & Bapco, E.L. (1984). *Professional women and minorities: A manpower data resource service* (5th ed.). Washington, DC: Scientific Manpower Commission.

Wakil, S.P., Siddique, C.M., & Wakil, F.A. (1981, November). Between two cultures: A study in socialization of children of immigrants. *Journal of Marriage and the Family,* 929-940.

Wright, E. (1979). *Class structure and income determination.* New York: Academic Press.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 462-473*

Marianne Gareau

and

Don Sawatzky
University of Alberta

Parents and Schools Working Together: A Qualitative Study of Parent-School Collaboration

The term parent-school collaboration *is a relatively recent one in education, and little research to date has attempted to study what parents and educators mean when they talk about parent-school collaboration. This article is a report of a study that focused on the perceptions of parents and educators of elementary schoolchildren. In-depth interviews were conducted with five participants: a school principal, a school counselor, a teacher, and two parents. Nine themes emerged from the analysis of these interviews; two were related to the importance of parent-school collaboration, and the remaining seven corresponded to important characteristics of parent-school collaboration. Issues surrounding these themes are discussed.*

L'expression collaboration parents-école *est tout à fait récente dans le domaine de l'éducation. Peu de recherche a été fait jusqu'à présent pour tenter d'expliquer ce que les enseignant(e)s et les parents veulent dire par la collaboration parents-école. Cet article rapporte une étude qui visait les perceptions des parents et des enseignant(e)s d'enfants scolaires du niveau primaire. Des entrevues détaillées ont été faites avec cinq participant(e)s dont un directeur d'école, un conseiller scolaire, une enseignante et deux parents (femmes). Neuf thèmes sont ressortis de l'analyse de ces entrevues dont deux étaient reliés à l'importance de la collaboration parents-école tandis que les sept autres thèmes correspondaient plutôt aux caractéristiques importantes de la collaboration parents-école. On discute d'autres questions reliées aux thèmes ressortis.*

It is generally agreed that good relationships between teachers and parents are important for schools, children, and families. However, the term *parent-school collaboration* is a relatively recent one in education, reflecting a general societal trend toward increased participation of parents in the educational decision-making process. Although the enhancement of parent-school collaboration is often touted as one of the priorities of education departments and school systems, very little research investigates what principals, counselors, teachers, and parents really mean when they talk about such collaboration. Until about 10 years ago, most literature on parent-school relationships focused on the term *parent involvement* (Sattes, 1985). Recently the more frequently used terms are *family-school partnerships* or *parent-school collaboration* (Coleman & Tabin,

Marianne Gareau is a doctoral student in counseling psychology. She has worked as a teacher and consultant with Edmonton Public Schools, and also spent seven years as a consultant for Alberta Education.
Don Sawatzky is a professor in the Department of Educational Psychology and has a special interest in family-school relationships.

1992; Epstein, 1992a). *Collaboration* is generally seen as a broader and more inclusive term than parent involvement because the latter term focuses mainly on the parents' role, whereas the former term focuses on the relationship between the home and the school and how parents and educators work together toward common goals.

The purpose of this study[1] was to understand how parents and educators describe parent-school collaboration, to learn how their views on this topic are similar, and also how they differ from one another. The investigation, conducted as part of the requirements for a master's degree in counseling psychology, employed in-depth, semistructured interviews, and grounded-theory techniques of analysis. The participants for the study included a school principal, a counselor, a teacher, and two parents.

This study was important for a number of reasons. In these times of social and economic challenges and change, it has become more important than ever for educators and parents to find more effective ways to work together for the benefit of children. In generations past, schools were set up as extensions of the family and society; in the past 30 or 40 years we seem to have lost the close relationship that once existed between parents and schools, and in many sectors an adversarial tone seems to have emerged between the school systems and the general public. Walberg (1984) quotes research that shows that the relationship between the home and the school has deteriorated in recent decades. Part of Alberta Education's (1994) Three-Year Business Plan for restructuring education involves increasing parental involvement in education. Another goal listed in the Three-Year Plan involves enabling parents and teachers to have "meaningful roles in decisions about policies, programs, budgets, and activities" (p. 9). In order for these goals to be realized, it is essential to begin by understanding what parents and teachers consider to be meaningful roles.

Current professional literature is filled with the concept of collaboration as a driving theme in education (Cook & Friend, 1991). Collaboration is discussed in terms of relationships in schools and school systems and among schools, families, and communities. Collaboration is defined in the literature in a number of different ways. Appley and Winder (1977) define collaboration as a relational system and suggest that competition and hierarchy no longer serve as an adequate value base for survival. Instead, they believe that we must come to view our environment as one that requires qualitatively new solutions based on an alternative value system, namely, what they call collaboration. In this new value system, caring (as opposed to competition or conflict) is central, and the relational system (rather than the individual) is the basic unit of collaborative effort. In a synthesis of research on organizational collaboration, Hord (1986) reviews the literature and concludes that "while there is little argument about the need for or value of collaboration ... there is disagreement about what 'counts' as collaboration" (p. 22). Hord also suggests that making a distinction between *cooperation* and *collaboration* can be useful because conflicts can arise when it is not clear which model is in process. In cooperation, parties agree only to work together; in collaboration, parties involved are also seen as sharing responsibility and authority for basic decision making. To clarify this distinction between cooperation and collaboration, Hord uses the metaphor of

the family where a mother cooperates with her son by allowing and encouraging his rock band to practice in their home, and the son cooperates with the mother by cleaning the house before guests arrive. On the other hand, the family collaborates in preparing a family meal together, and each offers some form of expertise that is rewarding to all.

Idol and West (1991) define collaboration more as "a structured process and an interactive relationship among individuals" (p. 72). Similarly, Moore and Littlejohn (1992) suggest that collaboration is "an interactive process in which all of the parties are equal" (p. 42), emphasizing the importance of equality of parents and educators. On the other hand, Cook and Friend (1991) refer to collaboration as a style rather than a process.

Historical and Theoretical Perspectives of School-family Relationships

There is no doubt that parents and families have always been involved in the education of their children, to the point that most writers refer to parents as being the oldest and most essential part of any education system (Berger, 1987). Parent involvement in the education of children has in fact been present since prehistoric times, with the family providing most of children's education informally until the introduction of public schooling. In the early 1900s the parent education movement began, and both parent education and involvement became institutionalized parts of the schools through parent-teacher associations. Until the 1940s, parent involvement in the schools, though indirect, was meaningful and constant (Comer, 1986). In small towns and rural areas parents had everyday contact with teachers. There was a sense of community and cultural uniformity, and the schools were natural extensions of the community. After World War II, however, came many technological, scientific, cultural, and social changes that resulted in a change in the level of trust and agreement between home and school. Training and achievement requirements were raised for educators and students alike; the content of the curriculum began to change, along with increased focus on teaching methods. By the 1950s the major responsibility for children's formal, academic education was delegated almost entirely to the schools (Topping, 1986). Schools discouraged parental intrusion, and the educative function of parents was downplayed.

Gradually the pendulum swung back once again, and the trend changed toward increased parent involvement in the schools. Comer and Haynes (1991) point out that the involvement of parents in their children's education is now widely accepted as desirable and even essential to effective schooling. However, they say that a significant number of educators are ambivalent about parent involvement in the schools, and parent participation is not significant in many schools even when parents are invited. Moles (1987), while promoting parent participation in education as an idea whose time has come, points out that the elements of confrontation and power sharing tend to make educators and school officials uneasy with parents in the roles of advocates and decision makers, roles where parents are becoming increasingly interested. Studies indicate (Lightfoot, 1978; Ost, 1988) that the majority of teachers and principals see the ideal relationship with parents as one in which parents support teacher practices and schools in general, carry out requests, but do not interfere with plans and decisions. Although parent involvement in instruction has been clearly linked to student success (Epstein, 1992a), teachers and administrators

often fail to establish strong links between home and school (Collinge & Coleman, 1992).

Recent Research on Parent-school Collaboration

A 1988 study by Leitch and Tangri on barriers to home-school collaboration focused on junior high schools and assessed both parents' and schools' concerns. The most frequent barrier as seen by the teachers was the parents, and their attitudes toward the school. When parents were asked to identify barriers to collaborating with the schools, they also centered the barriers in themselves, citing work responsibilities, lack of time, health problems, and economic differences between themselves and teachers. From this study, Leitch and Tangri (1988) concluded that the major barriers to collaboration are the lack of specific planning and the lack of knowledge about how teachers and parents can use each other more effectively.

Recent studies by Lindle (1989) and Lindle and Boyd (1991) asked parents to reflect on the worst and the best experiences they had with any schools and also asked teachers what they thought parent preferences were. Surprisingly, the preferences of parents were not what school personnel thought they were; according to this study parents viewed a professional, businesslike manner as undesirable and reported a personal touch as the most enhancing factor in school relations. All parents seemed to prefer relationships of a less formal nature with their child's teachers. Lindle (1989) concluded that parents do not want a professional-client relationship with the school: they want to be seen as equal partners with schools in the rearing of children.

Another group of researchers at Simon Fraser University have developed an intervention study that attempts to identify the attitudes among parents, teachers, and students that influence collaboration. They conclude that it must be teachers who permit parent-school collaboration (Coleman & Tabin, 1992); that is, they must realize the importance of teacher invitation, legitimizing, facilitating, and encouraging parent collaboration as part of their roles as teachers. However, Coleman and Tabin (1992) find little evidence that teachers are presently doing this; most classrooms are not characterized by strong positive relationships with parents. Coleman, Collinge, and Seifert (1992) reject the concept of parents as clients of the school, which they believe implies an arm's-length relationship. They also see the notion of parents as clients as inherently at odds with the notion of collaboration, which they believe is the more desirable relationship between schools and parents. Instead, they describe parents and students as participants in the school community and believe that "parents will have more influence, and will be more satisfied with the schools, if they are perceived as partners rather than as consumers or adversaries" (Coleman et al., 1992).

There is a definite trend in our society toward recognizing the rights of parents to make educational decisions concerning their children, along with a trend that emphasizes the importance of parents and educators working together to meet the educational needs of the children in their care. The present study was undertaken in an attempt to understand how parents and educators in Alberta describe the current reality in terms of parent-school collaboration. Furthermore, we wanted to learn more about the participants' hopes or vision for the future in terms of the relationship between parents and the schools.

Methodology

We chose a qualitative, human science research approach (Bogdan & Biklen, 1992; Patton, 1990; van Manen, 1990), because of its congruence with the topic and our personal beliefs about how we can best learn about another person's perspectives. In making a decision regarding the participants in this study, we were aware that the most basic qualification was that participants have "salient experiences of the phenomenon in their everyday worlds" (Becker, 1986, p. 105). Each of the participants in this study had direct experience of parent-school collaboration. Because one of the goals of the study was to explore the diversity of various perspectives on parent-school collaboration, we chose to interview a principal, a teacher, a counselor, and two parents. The decision to include two parents was made in order to provide some balance between the school/parent perspectives. Finally, we decided to select all the participants from one school in the hope that there would be some common ground in their experiences. Each of the participants was either a staff member of Holy Spirit Elementary School or had children who attended the school. Following is a brief description of each of the participants (names of individuals and the school have been changed to keep identities anonymous).

Stance, a principal. Stance has been the principal at Holy Spirit School for the past five years. He is married, has two children, and has been working as a teacher and administrator for over 20 years.

Bernard, a counselor. Bernard has been the school counselor at Holy Spirit for 17 years and has been working for the school district for 23 years. He is also married and has three children.

Clare, a grade 3 teacher who has been teaching elementary school for 20 years. She has also worked as a district office consultant for two years. She is married to a high school teacher and has three daughters.

Anne, a parent of two children (Julie, grade 5 and Justin, grade 3). Anne has a son in Clare's class this year, and Clare suggested her as a parent participant for the study. Anne is a nurse and a single parent and has raised both children on her own since they were very young.

Denae, a parent of three children (Veronique, grade 5, Melanie, grade 4, and Lisa, grade 1). Denae's name was suggested for this study by Bernard. Denae is a full-time homemaker, and her husband is a self-employed businessman.

Individual, in-depth semistructured interviews were used as the primary basis for the study in order to obtain as rich data as possible (Patton, 1990). An interview guide was followed, but participants were encouraged to digress and expand on their thoughts, and probing questions were used to elicit greater depth of responses. The opening item for each interview was: "Tell me about parents and schools working together." Other questions included asking participants how they knew when parent-school collaboration was occurring, asking them to describe a situation where they felt collaboration *had* occurred and another situation where it *had not* occurred, asking them to identify barriers to collaboration, to define or describe collaboration, and to tell about their vision of collaboration. I found that one of the later questions in the interview, "Is there anything else that I need to know about you personally in order to really understand your thoughts about this topic?" was the question that seemed to yield some of the richest data.

Each participant was interviewed twice, the first an open-ended interview, and the second a validation of my interpretations. Our intention during the interviews was to be as nondirective as possible, to avoid influencing the participants' responses. The first participant to be interviewed was Stance, the principal; followed by Bernard, the counselor; Clare, the teacher; and finally, Anne and Denae, the parents. With the exception of Stance, who was interviewed in late February 1994, all other participants were first interviewed during a two-week period in May 1994.

The data for the study consisted of the five taped interviews, the written transcriptions of these interviews, field notes, and journal reflections. In analyzing the data, each of the interview transcripts was first broken down into meaning units, and each unit was carefully paraphrased (first-level abstraction). Working with the help of a colleague, we then assigned a tag or label to each meaning unit (second-level abstraction). These tags were then clustered together into groups or topics for analysis, in order to develop the main themes for each participant (third-level abstraction). At this point, we returned to each participant for the second interview, in order to validate our analysis by seeing if each one agreed with it. After analyzing each interview individually, we began a comparative analysis. First we looked for themes that were specific to parents and specific to educators; second, we looked for some of the overall themes or issues (topics) that applied to all the interviews.

Findings

Nine major themes or issues emerged from the analysis of the interviews. The first two themes (children live in two worlds, and changes in society) address the question of the importance of parent-school collaboration, whereas the other seven themes (communication, trust/openness/honesty, positive/caring attitudes, personal connections, being equals, power/conflict/roles, and schoolwide commitment) can be seen as characteristics of that collaboration.

Why is parent-school collaboration so important?

Children live in two worlds

Each of the participants in the study was committed to the value and importance of parent-school collaboration. The most frequently mentioned reason for this was related to the "diffuse boundaries" for the child between the home and the school, or what Christenson, Rounds and Franklin (1992) refer to as the "mutually influencing quality" among home and school experiences. For example, Anne described parent-school collaboration as an ongoing relationship, where both parties have to work as partners because they both have a vested interest in the child:

> Not that they're the educators and I'm the parent, and when he comes home from school, that their job ends and mine begins, it's an ongoing relationship that happens, from day to day.

Clare, the teacher, also talked about being keenly aware of the connection between the home and the school for the child:

> I keep seeing, like, it isn't the child here, totally distinct, in this classroom, from his home life ... It's not, we carry them back and forth, the child carries [his

> experiences] back and forth ... We need to realize that the connection is so important.

Changes in society

The other theme that emerged when participants spoke about the importance of parent-school collaboration was that of change. Without exception, everyone spoke about his or her belief that we are in a time of change with respect to the roles of parents and schools. Some of the participants saw this change as connected with the current political situation in Alberta, with budgetary restrictions that may result in expecting parents to play a greater role in the governance and implementation of educational programs. Participants also spoke about how this generation of parents has more education, as well as higher expectations of the schools—they are "involved more, questioning more." Bernard, the counselor, talked about how he believes that collaboration between schools and parents has pretty much been kept at a "surface level" until quite recently: "We're only beginning to see what it really means for parents and teachers and other school personnel to really work closely together." He sees that parents are now more knowledgeable and are less willing to just accept the schools having total authority over educational decisions. Stance, the principal, also saw himself in process and changing some of his own views about the role of parents. He said that he realized that parents have not yet been brought in as full members of the educational team.

What are the characteristics of collaboration?

Communication

Both Clare, the teacher, and Anne, a parent, spoke about collaboration mainly in terms of the communication between the home and the school. All the participants believed that without communication, collaboration was impossible. However, for the parents and the teacher, communication was seen mainly in an individual, child-related sense; for Stance, the principal, a major concern was that parents and teachers communicate openly and honestly, as well as listening to various points of view. For both parents "being informed," "feeling listened to," and "having a voice" were among their strongest needs in their relationship with the school.

Trust, openness, honesty

Trust was talked about as an important characteristic in parent-school collaboration by all the participants. However, not everyone means the same thing when they talk about trust. For example, for both parents trusting the school meant knowing that the teachers would be able and willing to look after their children's needs and recognize each child's uniqueness. For Denae trust also meant that the school would keep her informed about what was going on and that it would not do things that were incompatible with the school's Christian mission. On the other hand, when the educators spoke about trust they were referring more to the importance of parents and educators having a basic belief in each other's good intention. For example, Stance would like parents to trust that the school is doing their best and that the educators are "professionally to know" what is in the child's best educational interest.

Each of the participants spoke about how trust and the mutual respect that ensues are developed gradually over time. There needs to be an openness and

honesty between people in order for this trust to develop; participants all believed that we can only reach the deeper levels of collaboration once these are established.

Positive, supportive, caring attitudes

Everyone spoke about the importance of parents and educators being positive and supportive toward one another. Clare talked about how, for her, collaboration has to be based on an environment of caring; Denae insisted that her comments must be viewed as supportive and positive, and not interpreted as critical. Stance said that he found that negative people "really get in the way of collaboration," and Bernard said that people must give each other positive recognition for what each has to offer.

Personal connections

In Lindle's (1989) research, establishing personal connections was seen by parents to be the most important characteristic of a collaborative relationship. This finding was also recognized as significant by both parents and educators in the present study. For Denae, knowing that the school people "really care" for the children is the most important thing. Stance talked about many of the informal contacts that he makes with parents, such as calling them at home to share some of the good things that their children have done at school. Clare, the teacher, establishes a relationship on a first-name basis with parents, makes face-to-face contact whenever possible, and talks about the bond that she feels with the parents of her students. Bernard, as the counselor, also emphasized the importance of caring for the parents as well as the students. In talking about this personal connection, the educators in this study felt that they were the ones with the greater responsibility for initiating this type of relationship.

Being equals

Each of the participants spoke about the importance of parents and educators seeing each other as equals in order for true collaboration to occur. The parents did not feel that this sense of equality always existed, although they recognized and appreciated it when it did. Bernard said he believed that "if we don't begin with a strong commitment to equality between parents and the school, the whole aspect of team is gone." For Stance, one of the ways for this to happen is for parents and educators to begin to "share roles a bit," for educators to give up some of their "professionalism" and to acknowledge that parents also have a lot to teach them and to contribute to the child's education.

Power, conflict, and roles

Differences in power between parents and educators is an issue that seems to be of concern mainly to the parents. Both parents talked of being worried about how the educators were perceiving them. They feel vulnerable in a way that seems qualitatively different for parents than it does for the educators. As Anne told me about some of her interactions with the school concerning her son Justin's program, she said, "There were times when I thought they were thinking, 'here she comes again!'" Similarly, Denae worried that Stance might misinterpret her statements at the parent meeting where she disagreed with him, and thought that he might think that she was not being supportive of him and of the school. Parents often feel powerless when they interact with school

personnel. Denae summed up some of her feelings by saying, "It's like it's a partnership as long as we all agree."

For the educators the concern is more related to what Clare refers to as the fine line regarding parents' power to make school decisions. She expresses concern about how much ownership parents should have for education. Stance also shared this concern, saying that he is "never sure at what point it's appropriate to make sure that there are safeguards in place." Stance says that although he recognizes that parents and educators share responsibilities, there are still some areas where the lines have been drawn, at least in an informal way. Stance admits that he finds this confusing and says that he "wrestles all the time" with trying to determine what the areas are where parents might have as "full a range of decision making" as teachers and administrators. Recognizing that we are in a time of changing roles and expectations, he says, "I'm just not sure where the parameters are right now, anymore."

Although everyone in this study recognized that conflict is inevitable, not everyone is comfortable in its presence. Denae said that she feels very uncomfortable with conflict and dissension, feeling that "dissension does not serve any purpose." Bernard, however, believes that if people are truly working together, conflict is inevitable, and that unless we begin to address the conflict areas, we will remain at what he calls a surface level in terms of collaboration. In talking about conflict Stance explained his belief that there needs to be at least a minimum level of basic agreement before collaboration between parents and the school can occur.

Schoolwide commitment

The importance of having a schoolwide commitment to collaborative relationships is one that all three educators in this study mentioned. As they said, collaboration must start at the school level; it is highly doubtful that educators who are not engaged in collaborative relationships among themselves will engage in collaborative relationships with parents. At Holy Spirit School there appears to be an atmosphere of working together that has also had an impact on parent-school relationships. As Clare said, "collaboration is not something that can happen in isolation; it needs to be a schoolwide approach to working with others." Stance believes that it is essential that all the movement toward collaboration not come from the administrators; he encourages parents to deal directly with their children's teachers whenever possible. The parents in the study also recognized the atmosphere that is established in the school, an "upbeat, positive one that extends to all the families that come to the school." All of the participants mentioned the crucial role that the administrators have in establishing this schoolwide commitment to collaborative relationships.

An Important Dilemma

One of the important dilemmas highlighted by this research is related to the difficulty of introducing a collaborative culture into what has traditionally been a hierarchical organization. It appears that school personnel experience feelings of vulnerability related to the potential loss of some of their professionalism. As long as school personnel relate the role of professional to that of expert, it will be difficult to achieve more than surface-level collaboration. In the way that both parents and educators referred to the issue of trust, for

example, it became evident that the school was strongly viewed as the body that would ultimately make educational decisions.

The focus of the parents in the study was to be able to trust the school to make these important educational decisions concerning their child. Both parents shared feelings of vulnerability vis-à-vis the school and concerns about how they might be viewed by the school and how this might in turn affect their children. This feeling of vulnerability was not expressed by the educators in the study, although they all shared an interest in hearing from parents in honest and open exchanges. All three educators in this study saw the ideal relationship with parents as going beyond parents being merely supportive of the school. At the same time, they expressed concerns about how far they could go in sharing their power with parents before they lost some essential autonomy as professionals. Educators here seem to be looking for some balance (Litwak & Meyer, 1974) and in this sense are sharing some of the concerns that emerge as the roles of parents in the education process are reexamined. Although we as educators verbalize definitions of collaboration, there are indications that we are still falling short of understanding and implementing true collaboration. The term collaboration has been discussed here in terms of the introduction of a new culture based on equal participation of parents and educators in educational decisions.

The dilemma exists for both parents and educators. For parents it entails relinquishing the security that comes when someone else has final responsibility for making educational decisions. For educators, what has to be relinquished is the notion of professionalism that is based on hierarchy and power.

The Role of the School Counselor

Epstein (1992b) notes that few leaders in the field of school-family relations have been school psychologists or counselors, in spite of the natural connections between their skills and interests and the needs of schools for better school and family connections. Over 10 years ago Lombana and Lombana (1982) suggested that if parent-school partnerships are to be strengthened in the face of opposing cultural, geographic, and attitudinal forces, school counselors should assume leadership roles in the school, particularly in developing and implementing plans for working with parents. They also stressed that these must be plans that meet the expressed parental needs, as well as ensure the most productive use of available resources. Noting that teachers and administrators are generally ill-prepared to work with parents, Lombana and Lombana (1982) advocated for counselors to take a more active role in the school in consulting with and supporting teachers in their work with parents.

These ideas proposed by Lombana and Lombana (1982) are closely related to those of Weiss and Edwards (1992), who suggest that school psychologists might assume the role of family-school coordinators; and Epstein (1992b) who advocates for the school psychologist to become the psychologist of the school. There have been some longstanding debates about ways for school psychologists to broaden their influence by working with all staff, students, and families (Epstein, 1992b) and to extend some of the special education practices to all families in a school. The results of this study would support these ideas and also give some direction to school psychologists who are interested in broadening their roles.

Conclusions

The term *parent-school collaboration* is a relatively recent one in education, reflecting a significant societal paradigm shift that is also occurring in other institutions such as health care and social services. Parents and schools have shared responsibility for children's education ever since public schools came into existence, but the recognition that parents might have equal rights in educational decision making is a concept that has emerged only over the past two or three decades (Cross, 1989). It is becoming an increasingly accepted fact that education, like other social services in our democratic society, is everybody's business; it is especially the business of those parents who have children in schools.

Note

1. This study was part of a master's thesis written by the first author.

References

Alberta Education. (1994). *Meeting the challenge: Three-year business plan*. Edmonton, AB: Author.

Appley, D.G., & Winder, A.E. (1977). An evolving definition of collaboration and some implications for the world of work. *Journal of Applied Behavioral Science, 13*, 279-290.

Becker, C.S. (1986). Interviewing in human science research. *Methods, 1*, 101-124.

Berger, E.H. (1987). *Parents as partners in education: The school and home working together* (2nd ed.). Toronto, ON: Merrill.

Bogdan, R.C., & Biklen, S.K. (1992). *Qualitative research for education*. Toronto, ON: Allyn and Bacon.

Christenson, S.L., Rounds, T., & Franklin, M.J. (1992). Home-school collaboration: Effects, issues, and opportunities. In S.L. Christenson & J.C. Conoley (Eds.), *Home-school collaboration: Enhancing children's academic and social competence*. Silver Spring, MD: National Association of School Psychologists.

Coleman, P., Collinge, J., & Seifert, T. (1992). *Seeking the levers of change: Participant attitudes and school improvement*. Paper presented at the Fifth Annual International Congress on School Effectiveness and Improvement, Victoria.

Coleman, P., & Tabin, Y. (1992). *The good teacher: A parent perspective*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Collinge, J., & Coleman, P. (1992). *Home-school collaboration: Teacher practices and parent attitudes*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco.

Comer, J.P. (1986). Parent participation in the schools. *Phi Delta Kappan, 67*, 442-446.

Comer, J.P., & Haynes, H.M. (1991). Parent involvement in schools: An ecological approach. *Elementary School Journal, 91*(3), 271-277.

Cook, L., & Friend, M. (1991). Principles for the practice of collaboration in schools. *Preventing School Failure, 35*(4), 6-9.

Cross, J. (1989). Recording: Parent, professionals and their perceptions of partnership. *Scottish Educational Review, 21*(2), 106-116.

Epstein, J.L. (1992a). School and family partnerships. In M.C. Alkin (Ed.), *Encyclopedia of educational research* (6th ed., pp. 1139-1151)), sponsored by the American Educational Research Association. New York: Macmillan.

Epstein, J.L. (1992b). School and family partnerships: Leadership roles for school psychologists. In S.L. Christenson & J.C. Conoley (Eds.), *Home-school collaboration: Enhancing children's academic and social competence* (pp. 499-515). Silver Spring, MD: National Association of School Psychologists.

Hord, S.M. (1986). A synthesis of research on organizational collaboration. *Educational Leadership, 43*(5), 22-26.

Idol, L., & West, J.F. (1991). Educational collaboration: A catalyst for effective schooling. *Intervention, 27*(2), 70-78.

Leitch, M.L., & Tangri, S.S. (1988). Barriers to home-school collaboration. *Educational Horizons, 66*(2), 70-74.

Lightfoot, S.L. (1978). *Worlds apart: Relationships between families and schools*. New York: Basic Books.

Lindle, J.C. (1989). What do parents want from principals and teachers? *Educational Leadership, 47*(2), 12-14.

Lindle, J., & Boyd, W.L. (1991). Parents, professionalism, and partnership in school-community relations. *International Journal of Educational Research, 15*(3/4), 323-337.

Litwak, E., & Meyer, H.J. (1974). *School, family and neighborhood: The theory and practice of school-community relations.* New York: Columbia University Press.

Lombana, J.H., & Lombana, A.E. (1982). The home-school partnership: A model for counselors. *Personnel and Guidance Journal, 61*(1), 35-39.

Moles, O.C. (1987). Who wants parent involvement? Interest, skills, and opportunities among parents and educators. *Education and Urban Society, 19*(2), 137-145.

Moore, D., & Littlejohn, W. (1992). Trends toward increased family-school collaboration. *Contemporary Education, 64*(1), 40-45.

Ost, D.H. (1988). Teacher-parent interactions: An effective school-community environment. *Educational Forum, 52*(2), 163-176.

Patton, M.Q. (1990). *Qualitative evaluation and research methods.* Newbury Park, CA: Sage.

Sattes, B.D. (1985). Parent involvement: A review of the literature. *Appalachia Educational Laboratory Occasional Paper Series.* Charleston, WV: Educational Services Office, Appalachia Educational Laboratory.

Topping, K. (1986). *Parents as educators: Training parents to teach their children.* London: Croom Helm.

van Manen, M. (1990). *Researching lived experience: Human science for an action sensitive pedagogy.* London, ON: University of Western Ontario.

Walberg, H.J. (1984). Families as partners in educational productivity. *Phi Delta Kappan, 65,* 397-400.

Weiss, H.M., & Edwards, M.E. (1992). The family-school collaboration project: Systemic interventions for school improvement. In S.L. Christenson & J.C. Conoley (Eds.), *Home-school collaboration: Enhancing children's academic and social competence.* Silver Spring, MD: National Association of School Psychologists.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 474-481*

G. Patrick O'Neill
Brock University

Authorship Patterns in AJER: Forty Years in the Making

This article examines authorship patterns in AJER *from 1955 to 1994. Data were collected on the number of single, co- and multiple authorships for each year. The years were, in turn, divided into eight five-year intervals. Frequencies and percentages were generated for each interval and chi-squares were computed between intervals and overall. In addition, author/article ratios were calculated for each year. Results showed increases in the number of co- and multiple authorships with corresponding decreases in the number of single authorships. However, the trend was broken in 1990. The author/article ratios fluctuated somewhat, but generally paralleled the shifts within intervals. It was tentatively concluded that education is not unlike other fields, that it too has experienced changes in authorship patterns. Interpretations are offered along with recommendations for further research.*

Cet article examine le nombre d'auteur(e)s qui contribuèrent à la revue AJER *entre 1955 et 1994. Les données ont été collectionnées pour chaque année en utilisant les articles écrits par un(e) auteur(e), par deux auteur(e)s, et par multiples auteur(e)s. On divisa ensuite les années en huit intervalles de cinq ans chacune. Les fréquences et les pourcentages ont été produits pour chaque intervalle et des χ^2 ont été calculés entre ces intervalles et sur la période entière. En plus, la proportion d'auteur(e)/article a été calculée pour chaque année. Les résultats indiquent une croissance d'articles écrits par deux ou multiples auteur(e)s et une décroissance d'articles écrits par un(e) auteur(e) seulement. Cependant, l'année 1990 déviede cette tendance. Les proportions auteur(e)/article fluctuaient quelque peu mais elles égalaient généralement les changements à l'intérieur des intervalles. On peut conclure de façon tentative que l'éducation ne diffère pas des autres domaines et qu'en éducation même retrouve-t-on des changements dans les nombres d'auteur(e)s qui écrivent des articles. Cet article offre des interprétation ainsi que des recommandations pour une recherche plus poussée.*

Introduction

Recently authorship patterns in journals have been explored in relation to de Solla Price's (1963) famous three predictions. He posited that single authorships would be extinct by 1980, that more than half of all articles published by 1980 would be multiple-authored, and that we would "move steadily toward an infinity of authors per paper" (p. 89). His predictions have been tracked in fields such as anthropology (Choi, 1988), business and economics (Petry & Kerr, 1982), counseling (Gladding, 1984; Strahan, 1982; Zook II, 1987), finance (Schweser, 1983), life sciences (Burman, 1982; Satyanarayana & Ratnakar, 1989), medicine (Burman, 1982; de Villiers, 1984; Farr, 1984; Powers, 1988; Rosenfeld, 1991; Satyanarayana & Ratnakar, 1989), nursing (Norris, 1993), physical education (Crase & Rasato, 1992), and psychology (Mendenhall & Higbee, 1982; Sacco & Milana, 1984). The studies themselves have varied in

G. Patrick O'Neill is a professor in the Graduate Department, Faculty of Education. His interests are in the areas of special education, teacher education, and higher education.

design, but all have come to the same conclusion. In short, there has been a significant increase over the years in coauthored (i.e., double, dual, joint) or multiple (i.e., three or more) authored articles, or both. The increases, though, have not always supported de Solla Price's (1963) predictions. For example, not one of the studies supports his first prediction; some, particularly in medicine, support his second prediction; and not one supports his third prediction.

Background

To date there have been only two authorship studies in education, neither of which has been Canadian. The first investigation (Lyle, 1986) looked at authorship patterns in volumes one through 22 (1963 to 1984) of the *Journal of Industrial Teacher Education*. He excluded editorials, book reviews, and other special sections from his analysis. Lyle found that the number of single authorships dropped below 70% in 1971 and have remained so for all but two years since. In contrast, the number of coauthorships increased from 7% in 1964 to a high of 44% in 1981. As well, coauthorships accounted for more than 20% of all authorships between 1971 and 1984. Multiple authorships showed a similar pattern, but remained substantially below that of joint authorships. They ranged from zero to a high of 16%. The author/article ratios reflected these changes being generally higher for the years 1971 to 1984.

The second inquiry (Swafford, 1990) examined authorship patterns in the *Journal of Educational Administration* from 1963 to 1987. All articles and research reports were included in the analysis. Excluded were editorials and book reviews. In order to detect trends, Swafford divided his analysis into five equal time periods as follows: 1963-1967, 1968-1972, 1973-1977, 1978-1982, and 1983-1987. He found that between 1963 and 1967 only 2% of the articles were coauthored. However, this figure rose to 18.2% for the 1968-1972 period and to 29.6% for the 1973-1977 period and has remained fairly steady since. Multiple authorships showed no discernible pattern as they accounted for fewer than 3% of all articles published between 1963 and 1987.

Problem

The purpose of this study is to look at authorship patterns in a Canadian education journal, namely, the *Alberta Journal of Educational Research* (*AJER*). Specifically, this study addresses the following six questions:

1. Overall, have coauthorships increased in *AJER*?
2. Overall, have multiple authorships increased in *AJER*?
3. If so, what pattern, if any, do the increases follow?
4. Is there a statistically significant relationship between the number of authors per article and dates of publication?
5. Overall, has there been an increase in *AJER*'s author/article ratio?
6. If so, what pattern, if any, do the ratios follow?

Procedures

AJER was chosen as the target journal because it is the oldest national refereed journal in education and one of the most highly respected in the country. All major articles including perspectives and symposia from issue 1, volume 1, 1955 to issue 4, volume 40, 1994, were included in the analysis. Excluded were editorials, book reviews, essay reviews, journal reviews, rejoinders and replies, research abstracts and notes, memorials, bibliographies, department reports,

book notices, conference notices, and advertisements. The data were grouped into eight five-year intervals as follows: 1955-1959, 1960-1964, 1965-1969, 1970-1974, 1975-1979, 1980-1984, 1985-1989, and 1990-1994. Each interval contained 20 issues.

The analysis was divided into two parts. The first addressed questions one to four; the second addressed questions five and six. Part one included the compilation of a record sheet for each of the eight intervals. There were three steps. First, the number of authors, that is, single, double, and multiple for each year and for the total interval was tabulated. Second, the number of authors per category was converted to percentages for each year and for the total interval. Third, chi-squares were calculated between each interval and overall. A 2 x 2 design was employed in which single authorships comprised one category and co- and multiple authorships comprised the other.

The second part of the analysis included the calculation of 40 author/article ratios. This was achieved by dividing the number of articles per year by the number of authors per year.

Findings

Table 1 summarizes a typical record sheet that was compiled for each interval. Included is the number of *AJER* authors per article for the years 1955 to 1959 inclusive. As illustrated, the period was dominated by single authorships. More than 85% of the authorships were single whereas joint authorships accounted for about 13% and multiple authorships accounted for about 2% of the total.

Table 2 gives the breakdown for the eight five-year intervals. Included are the totals for each period. As shown, single authorships accounted for more than 58% of all authorships between 1960 and 1974. However, the figure dropped to less than 50% between 1975 and 1989. At the same time, there was a corresponding rise in both joint and multiple authorships. In fact dual and multiple authorships accounted for more than 50% of all articles between 1975 and 1989. The biggest increase was in two-authored articles. Multiple authorships remained consistently below that of coauthorships, although they too increased sharply between 1980 and 1989.

Table 1
Percentage and Number of *AJER* Authors per Article for the Years 1955-1959

No. of Authors	1955		1956		1957		1958		1959		Totals	
	%	n	%	n	%	n	%	n	%	n	%	n
One	68.42	(13)	95.24	(20)	90.90	(20)	86.96	(20)	83.33	(20)	85.32	(93)
Two	26.32	(5)	4.76	(1)	4.55	(1)	13.04	(3)	16.67	(4)	12.85	(14)
Three or more	5.26	(1)	—	(0)	4.55	(1)	—	(0)	—	(0)	1.83	(2)
Totals	100.00	(19)	100.00	(21)	100.00	(22)	100.00	(23)	100.00	(24)	100.00	(109)

Table 2
Percentage and Number of AJER Authors per Article for the Eight Intervals Between 1955 and 1994

No. of authors	*1955-1959* %	n	*1960-1964* %	n	*1965-1969* %	n	*1970-1974* %	n	*1975-1979* %	n	*1980-1984* %	n	*1985-1989* %	n	*1990-1994* %	n
One	85.32	(93)	66.47	(84)	58.65	(78)	60.90	(95)	49.30	(71)	42.65	(58)	46.72	(57)	57.66	(79)
Two	12.85	(14)	33.33	(42)	35.34	(47)	30.77	(48)	44.45	(64)	42.65	(58)	38.52	(47)	29.20	(40)
Three or more	1.83	(2)	—	(0)	6.01	(8)	8.33	(13)	6.25	(9)	14.70	(20)	14.76	(18)	13.14	(18)
Totals	100	(109)	100	(126)	100	(133)	100	(156)	100	(144)	100	(136)	100	(122)	100	(137)
χ^2		6.64[3]		1.78		.51		4.07[2]		1.25		.43		3.10[1]		33.86[4]

[1] $p<.10$
[2] $p<.05$
[3] $p<.01$
[4] $p<.001$

The latest interval (1990-1994) proved to be somewhat of an anomaly. The number of multiple authorships remained fairly constant, but there was a shift in the number of double and single authorships.

The chi-square values between intervals and overall are given at the bottom of Table 2. The chi-square values were significant between the first two intervals (χ^2 (1, N = 235) = 6.65, p<.01), the fourth and fifth intervals (χ^2 (1, N = 300) = 4.07, p<.05), the seventh and eighth intervals (χ^2 (1, N = 259) = 3.10, p<.10), and overall (χ^2 (1, N = 1063) = 33.86, p<.001). These data would suggest that there is indeed a significant relationship between the number of authors per article and dates of publication. That is, the more recent the publication, the greater the number of authors. And although the relationship is not always significant, the overall is significant indicating that the direction is continuous, at least between 1955 and 1989.

Table 3 presents the author/article ratio for the years 1955 to 1994. The ratios fluctuated somewhat, but generally show marked increases, especially in 1960, 1965, 1968, 1970, 1974, 1980, 1984, 1986, and 1992. The lowest ratio was in 1956 (1.05); the highest was in 1992 (2.13). This means that there were 1.05 authors per article in 1956 compared with 2.13 authors per article in 1992.

Discussion

The findings of this inquiry are by and large similar to those of Lyle (1986) and Swafford (1990). All three studies found substantial increases in the number of co- and multiple authored articles although the numbers overall in *AJER* were somewhat higher. These results, then, would suggest that authorship differences, if any, between education and other fields are in degree, not pattern.

Evidence would also indicate that there may be plateaus in *AJER*'s authorship patterns. Plateaus are defined as, "spans of time in which the number of single, double and multiple authorships remains relatively stable" (Norris, 1993, p. 156). Accordingly, then, 1960 to 1974 and 1975 to 1989 would be two successive plateaus, each in turn having increased, but fairly consistent numbers of dual and multiple authored articles. This observation is supported in part by the fact that the chi-squares were significant at the beginning and end, but not within each of the periods.

If the trend were to continue, 1990 would be the beginning of another 15-year plateau, one where single authorships would again show marked decreases and joint and multiple authorships would in turn show marked increases. But this does not happen. Instead, there is a sudden reversal in the number of single and double authorships. And, if it were not for 1992, which had a disproportionately high number of dual authorships, the reversal overall would be far more pronounced.

This swing in direction would suggest that we may be entering an unexpected plateau, one where *AJER* is once more dominated by single authorships. Then, of course, the shift may be temporary, an aberration. In any case it is too early to predict the next plateau with any degree of accuracy because present data are limited to only five years and there are no comparison data. Future studies might address both these limitations.

Table 3
AJER Author/Article Ratio for the Years 1955-1994

Year	*1955*	*1956*	*1957*	*1958*	*1959*	*1960*	*1961*	*1962*	*1963*	*1964*
Articles	19	21	22	23	24	24	23	24	28	27
Authors	26	22	25	26	28	35	30	30	35	38
Ratio	1.37	1.05	1.14	1.13	1.17	1.46	1.30	1.25	1.25	1.41
Year	*1965*	*1966*	*1967*	*1968*	*1969*	*1970*	*1971*	*1972*	*1973*	*1974*
Articles	30	25	33	24	21	24	28	35	34	35
Authors	48	35	45	38	31	40	42	48	47	58
Ratio	1.60	1.40	1.36	1.58	1.48	1.67	1.50	1.37	1.38	1.66
Year	*1975*	*1976*	*1977*	*1978*	*1979*	*1980*	*1981*	*1982*	*1983*	*1984*
Articles	29	35	31	24	25	21	33	31	26	25
Authors	43	57	52	38	37	38	55	53	41	49
Ratio	1.48	1.63	1.68	1.58	1.48	1.81	1.67	1.71	1.58	1.96
Year	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*	*1992*	*1993*	*1994*
Articles	23	25	21	29	24	27	24	23	33	30
Authors	35	50	33	49	41	37	35	49	57	48
Ratio	1.52	2.00	1.57	1.69	1.71	1.37	1.46	2.13	1.73	1.63

Interpretation

Reasons for the increase in double and multiple authorships in *AJER* are speculative, but nonetheless worthy of note. I can think of at least three explanations; others may occur to the reader. First is the ever-increasing pressure on faculty to "publish or perish" (Day, 1994; Gladding, 1984; Mendenhall & Higbee, 1982; Sacco & Milana, 1984; Strahan, 1982; Zook II, 1987). This combined with an institutional reward structure that values single and joint authorships equally, as many do (Petry & Kerr, 1982), no doubt, helps account for the increased number of coauthorships.

Second, and closely allied, is the increase in collaborative efforts by graduate students and their supervisors. Indeed, the practice of mentoring has become more common (Mendenhall & Higbee, 1982; Zook II, 1987) for without "an established record of publication, Ph.D. holders will have little opportunity to obtain a teaching or research position in higher education" (Dill & Morrison, 1985, p. 177). I can identify with this factor, in particular, having coauthored an article in *AJER* with my advisor while pursuing a doctorate.

Third, perhaps, is the increasing complexity of research itself (Mendenhall & Higbee, 1982; Powers, 1988; Sacco & Milana, 1984; Zook II, 1987). The knowledge explosion combined with increased specialization and large, often interdisciplinary projects have forced researchers to pool their resources. In other words,

> As experimental methods have become more complex, and as complex statistical analyses have warranted a knowledge of computer applications, the need and likelihood for researchers to work together has increased. Researchers have divided the work according to each author's area of research expertise (e.g., subject matter, computer knowledge, methodology, statistics). (Mendenhall & Higbee, 1982, p. 6)

Conclusion

Clearly these results do not support de Solla Price's (1963) predictions. The number of authors per article has increased, but not to the extent proposed. For instance, the three-or-more category did not exceed 15% of all authorships in any one interval. This is a far cry from the 50% plus predicted by de Solla Price. In addition, there is no sign of a movement "toward an infinity of authors per paper." In fact, in the 40-year history of the journal, only four articles had five or more authors. The first appeared in volume 20, issue 3, 1974 and has six authors. The second appeared in volume 32, issue 1, 1986 and has five authors. The third and fourth articles appeared in volume 38, issue 1, 1992. The first has five authors; the second seven authors (the record holder).

Moreover, there may be a movement back to single authorships. Should this occur, we might safely conclude that de Solla Price's predictions were myopic; that, in reality, authorship patterns are cyclic, not linear.

References

Burman, K.D. (1982). Hanging from the masthead: Reflections on authorship. *Annals of Internal Medicine, 97*, 602-605.

Choi, J.M. (1988). An analysis of authorship in anthropology journals, 1963 and 1983. *Behavioral and Social Sciences Librarian, 6*, 85-94.

Crase, D., & Rascato, F.D. (1992). Single versus multiple authorship in professional journals. *Journal of Physical Education, Recreation and Dance, 63*, 28-31.

Day, R.A. (1994). *How to write and publish a scientific paper*. Phoenix, AZ: Oryx Press.

de Solla Price, D.J. (1963). *Little science big science*. New York: Columbia University Press.
de Villiers, F.P. (1984). Publish or perish—The growing trend towards multiple authorship. *SA Medical Journal, 66*, 882-883.
Dill, D.D., & Morrison, J.L. (1985). Ed.D. and Ph.D. research training in the field of higher education: A survey and a proposal. *Review of Higher Education, 8*, 169-186.
Farr, A.D. (1984). Multiple authorship of scientific papers. *Medical Laboratory Sciences, 41*, 1-2.
Gladding, S.T. (1984). Multiple authorship in the *Personnel and Guidance Journal*: A 12-year study. *Personnel and Guidance Journal, 62*, 628-630.
Lyle, S.P. (1986). Multiple authorship in volumes 1 through 22 of the *Journal of Industrial Teacher Education*. *Journal of Industrial Teacher Education, 24*, 105-108.
Mendenhall, M., & Higbee, K.L. (1982). *Recent trends in multiple authorship in psychology*. Paper presented at the Annual Meeting of the Western Psychological Association, Sacramento. (ERIC Document Reproduction Service No. ED 222 807)
Norris, R.P. (1993). Authorship patterns in *CJNR*: 1970-1991. *Scientometrics, 28*, 152-158.
Petry, G.H., & Kerr, H.S. (1982). Pressure to publish increases incidence of co-authorship. *Phi Delta Kappan, 63*, 495.
Powers, R.D. (1988). Multiple authorship, basic research, and other trends in the emergency medicine literature (1975 to 1986). *American Journal of Emergency Medicine, 6*, 647-650.
Rosenfeld, R.M. (1991). Clinical research in otolaryngology journals. *Archives of Otolaryngology—Head and Neck Surgery, 117*, 164-170.
Sacco, W.P., & Milana, S. (1984). Increase in number of authors per article in ten APA journals: 1960-1980. *Cognitive Therapy and Research, 8*, 77-84.
Satyanarayana, K., & Ratnakar, K.V. (1989). Authorship patterns in life sciences, preclinical basic and clinical research papers. *Scientometrics, 17*, 363-371.
Schweser, C. (1983). The economics of academic publishing. *Journal of Economic Education, 14*, 60-64.
Strahan, R.R. (1982). More on JCP publication: Single versus multiple authorship. *Journal of Counselling Psychology, 29*, 430-431.
Swafford, G.L. (1990). Window or mirror? A content analysis of the first 25 years of the *Journal of Educational Administration*. *Journal of Educational Administration, 28*, 5-23.
Zook II, A. (1987). Trend toward multiple authorship: Update and extension. *Journal of Counselling Psychology, 34*, 77-79.

The Alberta Journal of Educational Research *Vol. XLI, No. 4, December 1995, 482-488*

R.S. Brown

J.O. Anderson
University of Victoria.

and

M.M. McGee

Manifestation of Latent Thetas: A Comparison of Field-Study and Positivistic Approaches to the Investigation of Self-Nurturance

Self-nurturance is an essential trait of humans that appears to have its genesis in survival motivation (Gecko, 1985). It is thought that a balanced manifestation of self-nurturance leads to an independent yet meaningful interactive life in the human community (Rondolini, 1988a, 1988b, 1994). However, there is an absolute dearth of empirical studies into the phenomenon.

This article reports on the findings generated by a study into manifest self-nurturance. Two research approaches were used to collect and analyze data. A field study approach was used to observe and interview individuals who were self-described self-nurturants. The researchers shared the lives of a small, selected sample of individuals in order to explore the meanings and manifestations of self-nurturance. A positivistic approach was taken in another component of the study. This involved surveying a larger, cross-cultural sample of individuals and subjecting the data to analyses based on bivariate individual score curve theory (Rogers, McLean, & Hambleton, 1992). A final comparative analysis was planned in which the model generated from the field based study was tested using the empirical survey data in a procrustean, structural equation modeling approach.

The Field Study

The field study involved a prolonged, participatory observation of a purposely selected dyad of self-identified self-nurturants who had been living in Victoria, British Columbia for over a decade. This identification of self-nurturance was supported by both anecdotal descriptions generated from a conversation between researchers and the individuals of the dyad, and categorization of life-elements on the G-scale by a panel of trained raters (Gucci, 1982). The basic approach to data collection adopted was a multi-staged infiltration protocol (Brown, 1989). The initial engagement began with the senior author sharing the daily lives of the dyad. To minimize intrusive characteristics of the protocol, he

R.S. Brown and M.M.McGee are independent field researchers living in Victoria, British Columbia.
J.O. Anderson teaches in the Faculty of Education at the University of Victoria.

deliberately role-played a submissive in the social setting of the home. During this period, blatant recording of observations such as field notes, audio- or videotape could be not used unobtrusively; rather, recollective anecdotes were the prime data archived from the study. The participant observation continued over a three-year period and succeeded to the extent that the researcher effectively became a member of a triad of self-nurturants.

Concerns about the fidelity of the data coupled with the apparent success of the infiltration prompted the insertion of a second researcher into this social universe of self-nurturance. The junior author joined the ethnography in the third year at the invitation extended by the first researcher but supported by the original dyad. The staged infiltration resulted in two levels of anecdotal evidence allowing for a rich information base that could be fully investigated for strong patterns and evaluated by means of fold-over credence methodology (Traub, 1993).

The information collected over what has turned out to be a five-year period provides significant and consistent patterns of insight into the meaning and manifestations of self-nurturance that are reported in detail elsewhere (Anderson & Brown, 1994). It should be pointed out that the information collected suffered from what could best be termed restricted range. Given the social and psychological situation of the observed dyad and the reluctance of the researchers to impose direction on the inclinations and behaviors of the observed dyad, what has been collected provides an insider perspective on what could best be termed the indulgent extreme of self-nurturance. No one, not the observeds nor the observers, was willing to explore the lived experience of denial and privation. However, given the data that was collected (essentially the mid to extreme range of self-nurturance) projections can be made to the nihilist extreme.

For this article the three major themes are briefly described: bliss, rationalization, and the golden extreme. *Bliss* is the prime target state, the state of mind and soul of the individual after a period of growth and development induced by self-nurturant behaviors. A classic example of observed bliss occurred annually in the study dyad. Generally in mid-February when the cherry trees were blooming (the study was located in Victoria) the obvious self-nurturant behaviors of both individuals of the dyad included repeated comments regarding the "absolute beauty of the flowers and their elegant simplicity as individual flowers ... and voluptuous complexity when viewed in magnitude."

The dyad increased the frequency of watching the weather channel noting the temperatures and conditions on the prairies and *down east*. This frequency peaked when snowstorms and extremely low temperatures persisted, and this was accompanied by telephone calls to folks back home to see how things were. This was self-nurturant in the sense of contributing to the growth of satisfaction and adaptation to their life condition, thereby stabilizing the social, psychic, and physiological entropy of existence, in this way tending toward a state of bliss. The state of bliss was often accompanied by what some may consider trite comments, but to the individuals in a state of bliss were accurate, and perhaps profound, reflections of the phenomenon: "My, but isn't life grand" or "Ahh, this is wonderful."

Some (Maguire, 1976; Allard, 1985) have used the alternate term *smug* to label this particular manifestation, but we deemed that it did not capture the positive valencing attributed the state by the individuals who live the experience.

Rationalization was another strong theme in the information. Individuals often expressed reasons as to why self-nurturant behaviors were of merit. For example, an affinity for red wines of one of the participants was described in a growth-related manner in regard to the increased perceptivity and creativeness caused by the consumption of the beverage, and further, that recent articles in the popular press attributed enhanced health and longevity to the consumption of red wine. This was viewed as not only of personal benefit, but of general societal benefit not only in the contribution to the coffers of the federal treasury through exorbitant taxation on the alcohol, but also in the reduced drain on the medicare system over the long term. Such protracted, convoluted, and yet genuinely believed positions were held for most self-nurturant behaviors recognized by the individuals of the dyad.

The golden extreme was a pattern derived from the observational data rather than any text collected from conversation. If the self-nurturance involved consumption, red wine consumption for example, moderation would occur only if some external environmental constraints were imposed such as a lack of funds to obtain sufficient quantities. Otherwise consumption would be maximized. Another example occurred with the use of meditation that was viewed as self-nurturant in that mental restoration and invigoration, reestablishment of psychic center, and general rebalancing of the self could be attained. Both individuals practiced meditation to the point of prolonged catharsis, and in fact preferred it as a complete lifestyle. The only reason they maintained active professional lives was the need for revenue to provide what they viewed as necessary resources.

The Cross-cultural Survey

A survey instrument developed in a previous study (McGee, 1979) to measure the levels of general life satisfaction and self-nurturance (GLSSN) was administered to samples of individuals from the West Coast, the Prairies, central Ontario, Quebec, and the Maritimes. These populations were chosen to maximize variation. The GLSSN is a forced-choice, polychotomously scored, computer-mediated instrument that can be administered either individually using the specially developed GLSSN headgear or in groups by means of a Jumbotron. Demographics (home location, sex, age, and height) were also collected.

The GLSSN yields two scores for each individual: General Life Satisfaction (GLS) and Level of Self-Nurturance (LSN). Analysis included calculation of descriptive statistics and correlations and bivariate individual score curve theory analysis (BISCT). BISCT (Rogers et al., 1992) is essentially an inverted item response theory (IRT: Lord, 1980) approach. IRT characterizes item performance on the basis of person responses, whereas BISCT characterizes people on the basis of test scores. BISCT generates theta values that are descriptive of the bivariate status of an individual on the full range of the satisfaction/self-nurturant plane.

The first round of analysis was conducted on each sample separately and results were compared across samples. No significant differences were detected, so for further analyses data were pooled.

Descriptive statistics of the GLSSN were best summarized as a frequency distribution for the total sample of respondents (Figure 1). The Self-Nurturant scores were remetricated utilizing the nosological standardization algorithm of Fossey (1948) thus resulting in a scale of 0.0 to 1.0. From this distribution it can be seen that self-nurturance is essentially normally distributed within the pooled data. The average value was 0.62 with a standard deviation of 0.28. The extremes of the distribution were labeled deprivation and indulgence for what should be obvious reasons, the center could best be termed balanced self-nurturance (BSN).

The distribution of these data was not unexpected given the recent findings and opinions of Klien (1993). However, the empirical nature of these data certainly lends credence to the notion of the inherent ability of Canadians to look after themselves. Of greater import and uniqueness are the results of the BISCT analyses that yielded Self-Nurturant/Satisfaction plots (SNSPs). From these plots it became evident that within the pool of respondents were archetypical profiles that were more interpretable than simple distributions of pooled data, and were more amenable to the planned procrustean merger (Heitz, 1987) with the observation data of the field studies. Each of the found patterns is based on data from at least 300 respondents, and the resulting characteristic person curve (CPC) was smoothed using the LOWESS procedures (Wilkinson, 1990). It was believed that these merged data would then lead to the development of the structural model of self-nurturance.

The SNSP's resulted in highly interpretable representations of the data that resonated with the observed life of self-nurturants investigated in the field study component of these researches. On the basis of the research literature, a number of major patterns had been anticipated and were indeed found. The first major pattern anticipated was what we termed the *golden mean* (Figure 2a),

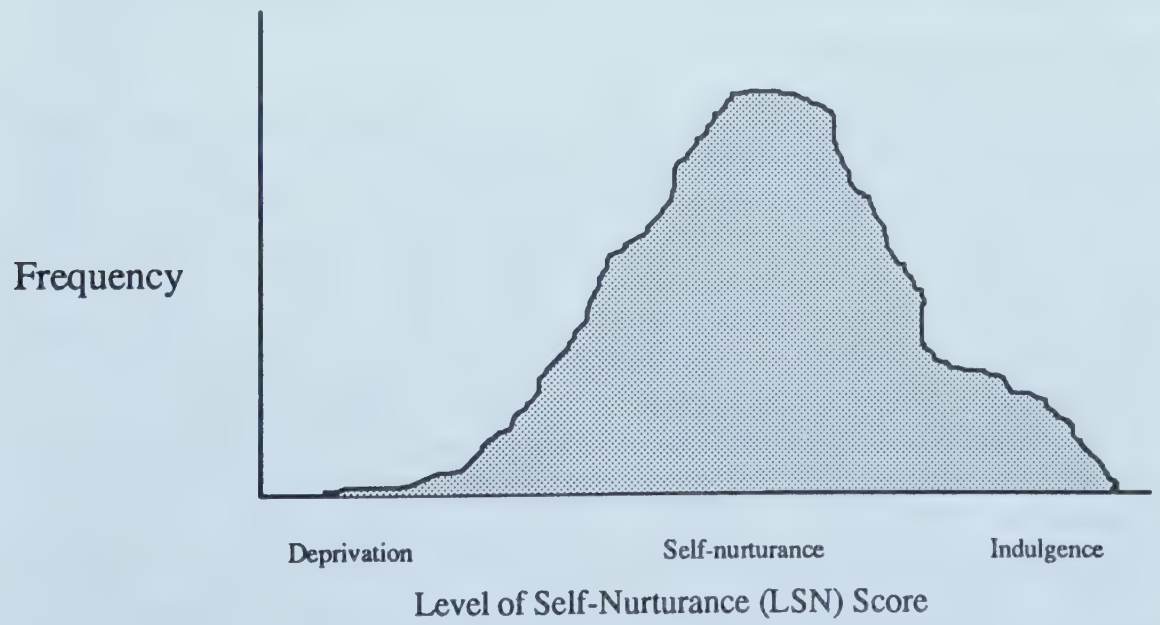


Figure 1. Distribution of self-nurturance scores in total sample of respondents (n=7,877).

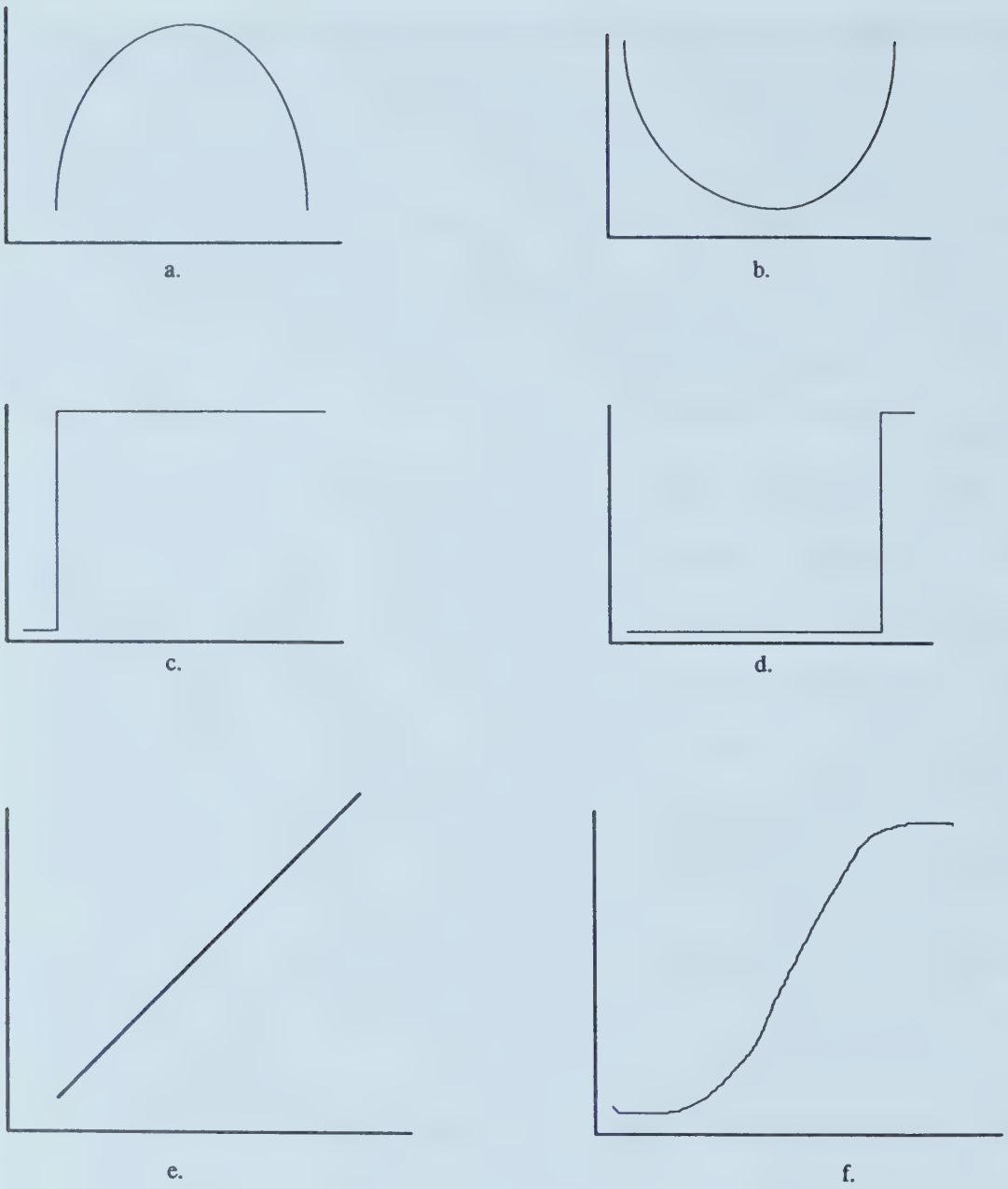


Figure 2. Person characteristic curves.

too much (or little) of a good thing is not all that satisfying. Rather, just enough is just right. Another pattern that was initially believed to be an artifact of data inversion but on further analysis proved to be accurate was the *guilt-ridden-masochistic indulgent* (Figure 2b) who is characterized as being most satisfied at the extremes of self-nurturance. From some of the available observational information, the basis to the lack of satisfaction with moderate levels of self-nurturance is associated with voluminous guilt that apparently dissipates as the state of indulgence is approached or an aesthetic privation is attained.

Departures from the continuous relationships to a dichotomous condition were also found (predominantly with data generated from administrators and accountants). The profile we labeled the *pollyanna* (Figure 2c) indicates an individual who is maximally satisfied with minimal levels of self-nurturance. The *sulky hedonist* (Figure 2d) on the other hand, was maximally dissatisfied with all but the indulgent levels of self-nurturance. Incidentally, the dichots

were generally reported to be extremely difficult individuals to live with given that they were either insufferably satisfied or unsatisfied.

The pattern associated with the majority of respondents, particularly those from the ranks of academe, shows a constantly increasing satisfaction with increasing levels of self-nurturance, the *material realist* (Figure 2e). The material realist simply is more satisfied with more self-nurturance, apparently without limit according to our data. A variation on this pattern is a *modulated material realist* (Figure 2f), which is characterized by limits to the change in satisfaction at the extremes of self-nurturance. The modulation is more characteristic of southern Ontario, whereas the nonmodulated material realist is more common among residents of the West Coast.

Some exploratory analyses conducted on small ($n<100$) datasets revealed some potentially significant variations. Figure 3 demonstrates what are termed guilt-dips in the otherwise linear relationship of satisfaction to self-nurturance in the material realist. It appears that some mr's actually have episodic lapses in increasing satisfaction as self-nurturance increases, a pattern we are confident is not associated with anyone reading this article.

Conclusion

It is obvious that the findings of both the field study and the survey allow for profound understandings of the human condition. It was the plan to synthesize the two datasets to develop and test a structural model of self-nurturance, and some work has proceeded in this direction. However, as this article was in preparation the field team defected to their study dyad, in effect creating a self-indulgent quadrad that has rejected the notion of research as a valued activity. In doing so, they absconded with all existing field notes and associated data. This results in an abandonment of the third stage of the project: the procrustean fitting of the empirical GLSSN data with the model of self-nurturance developed from the insights and meanings generated from the dyad of the field study. A preliminary model and novel display protocol had been developed but not yet fully evaluated. It is a sad note on which to end but it does accurately reflect the dangers of delving into the heart of human self-nurturance.

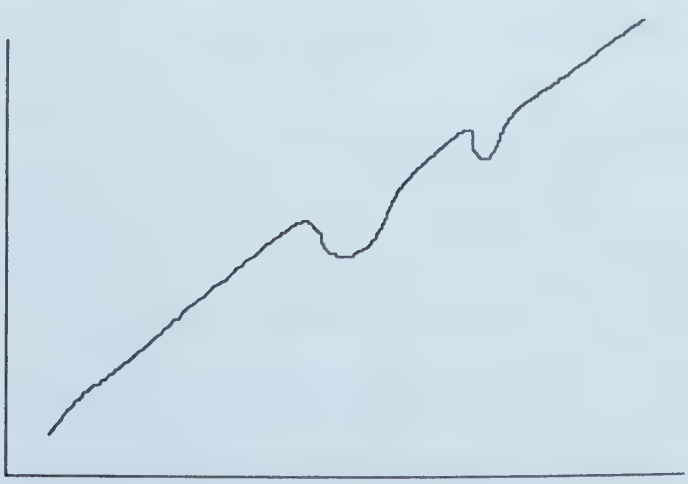


Figure 3. Person characteristic curves with guilt-dips.

References

Allard, M. (1985). *Bliss my soul what's wrong with me?* Memphis, TN: Hound Dog Press.

Anderson, J.O., & Brown, R.S. (1994). *Tending the personal garden.* Toronto, ON: Harrowsmith.

Brown, R.S. (1989). Aiming for non-obtrusiveness in observation. *Methodology in Observation, 25*(4), 233-235.

Fossey, D. (1948). Cross-specific standardization in primates. *Agronomist, 46*(1), 23-27.

Gecko, G. (1985). *Greed is good.* New York: Simian Press

Gucci, A. (1982). *Evaluating nurturance of individuals and groups.* Milan: Moderna Press.

Heitz, H.H. (1987). Blending qualitative and quantitative data meaningfully. *Journal of Applied Qualitative Analysis, 2*(2), 134-149.

Klien, R. (1993). The suppression of self-nurturance in academe. *Western Reports, 11*(6), 344-348.

Lord, F. (1980). *Applications of item response theory to practical testing problems.* New York: Erhlbaum.

Maguire, T.O. (1983). *Individualism as random rogue variables.* Millet, AB: Populist Press.

McGee, Q. (1979). Two essential dimensions of life: Satisfaction and self-nurturance. *WestCoast Life, 12*(3), 125-133.

Rogers, W.T., McLean, L.D., & Hambleton, R. (1992). *Individual score theory applications.* Denver, CO: Mile High Books.

Rondolini, H.L. (1988a). Self-nurturance and student achievement. *Journal of Educational Measurement, 24*(3), 345-356.

Rondolini, H.L. (1988b). Predictors of optimal self-nurturance: A factor analytic approach. *Multivariate Behavioral Psychology, 12*(4), 545-560.

Rondolini, H.L. (1994). Meanings of self-nurturance in the developing self. *Annals of Lived Hermeneutic Interpretations, 3*(1), 33-103.

Traub, R. (1993). Fold-over credence checks in multi-level information sets. *Journal of Applied Qualitative Analysis, 7*(2), 234-239.

Wilkinson, L. (1990). *SYSTAT.* Evanston, IL: SYSTAT Inc.

Index, Volume XLI, 1995

Amadeo Marquez, S. *The Spurious Relationship Between IQ and Social Behavior: Ethnic Abuse, Gender Ignorance, and Confounded Education*, No. 3, 277-282.

Anderson, J.O. *Manifestation of Latent Thetas: A Comparison of Field-Study and Positivistic Approaches to the Investigation of Self-Nurturance*, No. 4, 482-488.

Andrews, J.J.W. *A Classroom-based Social Skills Intervention for Children with Learning Disabilities*, No. 1, 84-102.

Barnett-Foster, D. *A Comparison of Undergraduate Test Response Strategies for Multiple-choice and Constructed-response Questions*, No. 1, 18-35.

Barrow, R. *Keep them Bells a-Tolling*, No. 3, 289-296.

Bateson, D.J. *How Can* The Bell Curve *Be Taken Seriously?* No. 3, 271-273.

Belke, T.W. *A Synopsis of Herrnstein and Murray's* The Bell Curve: Intelligence and Class Structure in American Life *(New York, Free Press, 1994)*, No. 3, 238-256.

Beynon, J. *Access and Aspirations: Careers in Teaching as seen by Canadian Univesity Students of Chinese and Punjabi-Sikh Ancestry*, No. 4, 435-361.

Biggs, J. *Assessing for Learning: Some Dimensions Underlying New Approaches to Educational Assessment*, No. 1, 1-17.

Bountrogianni, M. *Teachers' Views of Parents: Family Decision Making Styles and Teacher-Parent Agreement Regarding Homework Practices and Values*, No. 2, 175-187.

Brannigan, A. *IQ and Crime: Dull Behavior and/or Misspecified Theory?* No. 3, 316-321.

Brown, R.S. *Manifestation of Latent Thetas: A Comparison of Field-Study and Positivistic Approaches to the Investigation of Self-Nurturance*, No. 4, 482-488.

Cameron, J. *Introduction to Special Issue: Canadian Perspectives of* The Bell Curve, No 3, 237-237.

Cameron, J. *From the Editor*, No. 4, 377-377.

Connors, J.B. *Two Tails are Better than One: The Logic of* The Bell Curve, No. 3, 342-348.

Conte, R. *A Classroom-based Social Skills Intervention for Children with Learning Disabilities*, No. 1, 84-102.

Cousins, J.B. *Giving and Receiving Explanations in Cooperative Learning Groups*, No. 1, 103-121.

da Costa, J.L. *Teacher Collaboration: A Comparison of Four Strategies*, No. 4, 407-420.

Derwing, T. Review of *Genie: An Abused Child's Flight from Silence* by Russ Rymer, No. 1, 122-124.

Dolmage, R.W. *Accusations of Teacher Sexual Abuse of Students in Ontario Schools: Some Preliminary Findings*, No. 2, 127-144.

Epling, W.F. *The Bell Bottom*, No. 3, 302-303.

Fancher, R.E. The Bell Curve *on Separated Twins*, No. 3, 265-270.

Filipovich. T. *Teachers' Views of Parents: Family Decision Making Styles and Teacher-Parent Agreement Regarding Homework Practices and Values*, No. 2, 175-187.

Gambell, T.J. *Ethnography as Veneration*, No. 2, 162-174.

Gareau, M. *Parents and Schools Working Together: A Qualitative Study of Parent-School Collaboration*, No. 4, 462-473.

Gartrell, J. *The Spurious Relationship Between IQ and Social Behavior: Ethnic Abuse, Gender Ignorance, and Confounded Education*, No. 3, 277-282.

Gartrell, J. *University Performance of Nonmatriculated Admissions*, No. 4, 421-434.
Gibbs Van Brunschot, E. *IQ and Crime: Dull Behavior and/or Misspecified Theory?* No. 3, 316-321.
Golec, J. *University Performance of Nonmatriculated Admissions*, No. 4, 421-434.
Greene, M.L. *The Forgotten Link: Transitions from Graduate School to Classroom Teaching*, No. 2, 213-230.
Holdaway, E.A. *Expectations Held for Department Heads in Postsecondary Institutions*, No. 2, 188-212.
Hull, J. *Indian Control and the Delivery of Special Education Services to Students in Band-operated Schools in Manitoba*, No. 1, 36-62.
Hutton, G. *A Classroom-based Social Skills Intervention for Children with Learning Disabilities*, No. 1, 84-102.
Jones, D.R. *Expectations Held for Department Heads in Postsecondary Institutions*, No. 2, 188-212.
Kavanagh, M. The Bell Curve: *Rising Fears, Fallen Dreams*, No. 3, 297-301.
Kirby, J.R. *Intelligence and Social Policy*, No. 3, 322-334.
Krishnan, P. The Bell Curve: *Some Statistical Concerns*, No. 3, 274-276.
Krull, C.D. *IQ Testing in America: A Victim of its Own Success*, No. 3, 349-354.
Livingstone, D.W. *For Whom* The Bell Curve *Toils*, No. 3, 335-341.
Loomer, M. *A Classroom-based Social Skills Intervention for Children with Learning Disabilities*, No. 1, 84-102.
MacPherson, E.D. The Time Machine: *The Next Generation*, No. 3, 304-307.
Malicky, G.V. *Perceptions of Literacy and Adult Literacy Programs*, No. 1, 63-83.
McGee, M.M. *Manifestation of Latent Thetas: A Comparison of Field-Study and Positivistic Approaches to the Investigation of Self-Nurturance*, No. 4, 482-488.
Nagy, P. *A Comparison of Undergraduate Test Response Strategies for Multiple-choice and Constructed-response Questions*, No. 1, 18-35.
Norman, C.A. *Perceptions of Literacy and Adult Literacy Programs*, No. 1, 63-83.
Norris, S.P. *Format Effects on Critical Thinking Test Performance*, No. 4, 378-406.
O'Neill, G.P. *Authorship Patterns in AJER: Forty Years in the Making*, No. 4, 474-481.
Phillips, R. *Indian Control and the Delivery of Special Education Services to Students in Band-operated Schools in Manitoba*, No. 1, 36-62.
Pierce, W.D. *IQ Testing in America: A Victim of its Own Success*, No. 3, 349-354.
Polyzoi, E. *Indian Control and the Delivery of Special Education Services to Students in Band-operated Schools in Manitoba*, No. 1, 36-62.
Pratt, M.W. *Teachers' Views of Parents: Family Decision Making Styles and Teacher-Parent Agreement Regarding Homework Practices and Values*, No. 2, 175-187.
Priegert Coulter, R.P. *And Academic Sexism Too: A Comment on* The Bell Curve, No. 3, 308-311.
Purvis, C.R. *The Forgotten Link: Transitions from Graduate School to Classroom Teaching*, No. 2, 213-230.
Reviere, R. *"Slow Through The Dark": A Commentary on* The Bell Curve*: Intelligence and Class Structure in American Life*, No. 3, 312-315.
Robeck, E.C. *Distorted Vision: Education As Seen Through* The Bell Curve, No. 3, 367-376.
Robitaille, D.F. *Distorted Vision: Education As Seen Through* The Bell Curve, No. 3, 367-376.

Ross, J.A. *Giving and Receiving Explanations in Cooperative Learning Groups*, No. 1, 103-121.
Sawatzky, D. *Parents and Schools Working Together: A Qualitative Study of Parent-School Collaboration*, No. 4, 462-473.
Schleifer, M. *Should we Change our Views About Early Childhood Education?* No. 3, 355-359.
Siegel, L.S. *Does the IQ God Exist?* No. 3, 283-288.
Stenlund, K.V. *Teacher Perceptions Across Cultures: The Impact of Students on Teacher Enthusiasm and Discouragement in a Cross-cultural Context*, No. 2, 145-161.
Sveinson, W.J. *University Performance of Nonmatriculated Admissions*, No. 4, 421-434.
Symbaluk, D.G. *Book Review: Women's Work is Never Done: Resolving Gender Inequalities in Education*, No. 2, 231-235.
Toohey, K. *Access and Aspirations: Careers in Teaching as seen by Canadian University Students of Chinese and Punjabi-Sikh Ancestry*, No. 4, 435-361.
Wahlsten, D. *Increasing the Raw Intelligence of a Nation is Constrained by Ignorance, Not its Citizens' Genes*, No. 3, 257-264.
Wangler, D.G. *Is* The Bell Curve *A Ringer?* 360-366.
Wolbring, G. Review of *Violence and Abuse in the Lives of People With Disabilities* by Dick Sobsey, No. 1, 124-126.

For Contributors

Please prepare manuscripts in accordance with these guidelines.

Format

Manuscript in triplicate.
Typewritten or printed, double-spaced throughout.
Abstract of approximately 100 words on a separate sheet.
Manuscript not to exceed 6,500 words excluding graphics.
Author's name and affiliation on a separate title page; title only on first page of manuscript to ensure anonymity in the review process.

Referencing

Sources in parentheses after each reference (direct or otherwise), giving author's name (unless mentioned in text), year of publication, and page number(s) for direct quotes.
Direct quotes of 40 words or less in double quotation marks incorporated into text; those longer than 40 words in indented block format. Page numbers must be given.
All sources listed alphabetically at the end of the manuscript under the heading *References* using the APA style.
Explanatory notes, numbered consecutively and marked in the text with superscript numerals, may appear before the References under the heading *Notes*; any citations in notes follow the same format as other references.

Graphics

Each table and figure on a separate page at the end of the manuscript.
Figures, charts, and diagrams in camera-ready form.
Number tables in Arabic numerals and center over each heading, e.g., *Table 1.*
Number figures in Arabic numerals before the caption at foot of each graph, e.g., *Figure 1. ...*
Indicate placement of figures and tables in text, i.e., *Insert Table 3 here.*

Style

AJER's editorial style conforms closely to the *Publication Manual of the American Psychological Association.*
Editorial changes may be made to manuscripts.
For spelling, consult *Webster's New Collegiate Dictionary;* spelling in quoted material remains as in the original.

Diskettes

If manuscript is accepted for publication please supply a diskette from a Macintosh, IBM, or IBM compatible system. Microsoft Word and WordPerfect are the preferred software.

Submission to *AJER* implies the manuscript is not being considered for publication elsewhere.

AJER is indexed in the *Alberta Education Index, Canadian Education Index, Current Contents/Social and Behavioral Sciences, ERIC/Current Index to Journals in Education,* and *Social Science Citation Index;* appropriate articles are abstracted in *Cognitive Development Abstracts, Educational Administration Abstracts, Educational Technology Abstracts, Family Abstracts, Language Behavior Abstracts, Multicultural Education Abstracts, Psychological Abstracts, Research into Higher Education Abstracts, School Organization and Management Abstracts,* and *Sociology of Education Abstracts,* and reviewed by the *Journal for Research in Mathematics Education.*

If undelivered return to:

The Alberta Journal of Educational Research
845 Education Centre South
University of Alberta
Edmonton, AB
T6G 2G5

Publications Mail Registration Number 1438